Epidemiological and genetic associations between Cannabis Use Disorder and Major Depressive Disorder

Smolkina, Milana

Awarding institution:
King's College London

Epidemiological and genetic associations between Cannabis Use Disorder and Major Depressive Disorder

Milana Smolkina (BSc, MSc)

Addictions Department
Institute of Psychiatry, Psychology and Neuroscience
King's College London

A thesis submitted to King's College London for the degree of Doctor of Philosophy (PhD)

Abstract

Background: Cannabis is the most commonly used illicit drug in the United Kingdom and worldwide. It is associated with a number of negative outcomes, which includes developing Cannabis Use Disorder (CUD). Individuals who meet criteria for CUD are at heightened risk for experiencing Major Depressive Disorder (MDD), the leading cause of disability worldwide. While this association has frequently been reported, the underlying mechanisms remain controversial.

Aims of thesis: This thesis aims to investigate the degree of co-morbidity between lifetime rates of CUD and MDD, test whether this co-morbidity is accounted for by shared covariates, and test different twin models to investigate the sources (environmental or genetic) of and mechanisms underlying this co-morbidity.

Methods: Data analysis was conducted on a sample of 3824 Australian twins and their non-twin siblings. Epidemiological analyses, using multivariable logistic regressions, tested whether CUD and MDD were significantly co-morbid in this sample, and to what extent covariates influenced this relationship. Twin models – bivariate correlated liabilities, discordant twin and co-morbidity models – examined whether the co-morbidity between the disorders could be explained by a) shared genetic and environmental factors, b) causal processes, and c) 13 different models of co-morbidity.

Results: The epidemiological analyses found that MDD and CUD were significantly co-morbid in this sample: meeting diagnostic criteria for one disorder more than doubled the odds of meeting criteria for the other (odds ratio = 2.23, 95% confidence interval = 1.84–2.70). This co-morbidity could not be fully attributed to various psychiatric, trauma-related, parental, peer and demographic covariates. Bivariate twin analyses found that – when separated into genetic and environmental correlations – the only significant correlation between MDD and CUD was genetic (r =.41, 95% confidence interval = .24–.60). A possible causal relationship could not be excluded, because MDD and CUD were significantly associated (odds ratio = 2.83, 95% confidence interval = 1.12–7.19) in monozygotic twins discordant for both disorders. Co-morbidity model analyses indicated that the direction of influence was

from CUD to MDD, and that CUD risk factors may cause MDD symptoms, particularly in individuals at high risk of CUD.

Acknowledgements

I would like to thank my father for inspiring me to run mental marathons, Tanya for teaching me how to endure them, and my mother for making me feel at peace with failure.

Also, I would like to thank my supervisors Michael Lynskey, Katherine Morley, and Frühling Rijsdijk. They have been incredibly kind, patient and generous with their time for as long as I have had the pleasure of knowing them.

Publication

A published version of Chapter 6 can be found in:

Smolkina, M., Morley, K. I., Rijsdijk, F., Agrawal, A., Bergin, J. E., Nelson, E. C., … Lynskey, M. T. (2017). Cannabis and Depression: A Twin Model Approach to Co-morbidity. *Behavior Genetics, 47*(4), 394–404. http://doi.org/10.1007/s10519-017-9848-0

Table of contents

Table of tables

Table of figures

1 Introduction

1.1 Psychiatric co-morbidity: Cannabis Use Disorder and Major Depressive Disorder

The main diagnostic frameworks for mental health problems place disorders into clearly separated categories. This principle applies to both the International Classification of Diseases (ICD; World Health Organization, 1992) and the Diagnostic and Statistical Manual of Mental Disorders (DSM; American Psychiatric Association, 2013). Nevertheless, the co-morbidity – or co-occurrence of two or more psychiatric disorders within one individual – is highly frequent (Boyd et al., 1984; Kessler et al., 1994). The term 'co-morbidity' was initially introduced in a paper by Feinstein (1970). The definition includes disorders which occur at the same time and those which occur at different times throughout an individual's life. Feinstein (1970) recognised the importance of increased awareness of co-occurring conditions in clinical practice because individuals with co-occurring disorders differ from individuals with single disorders in multiple ways.

Psychiatric co-morbidity is associated with various negative life outcomes (Compton, Thomas, Stinson, & Grant, 2007; Swendsen & Merikangas, 2000). The co-morbidity of substance use problems and other mental health disorders has received particular attention because it is highly prevalent, and co-morbid individuals exhibit symptoms which are often more severe, persistent and difficult to treat (Kessler, 2004; Swendsen et al., 2010; Torrens, Mestre-Pintó, & Domingo-Salvany, 2015). According to a recent report by the European Monitoring Centre for Drugs and Drug Addiction (EMCDDA), individuals with a substance abuse problem co-morbid with an additional mental health disorder are at higher risk of suicide, are more likely to be unemployed or homeless, engage in behaviours which endanger their personal safety, and are more likely to be admitted to a hospital for emergencies (Torrens et al., 2015). Co-morbid individuals are also more likely to engage in violent or criminal behaviour, when compared to individuals with a single disorder (Torrens et al., 2015). Aside from the burden on the individual, societal costs related to mental health conditions co-morbid with substance use disorders are substantial (see Whiteford et al., 2013).

Unsurprisingly, identifying the factors which underlie co-morbidity is part of the 'Mental Health Research Priorities for Europe' established in 2015 (Wykes et al., 2015). This is one of the areas of mental health in which research investment is likely to lead to high returns. In 2010, mental disorders were conservatively estimated to cost €461 billion a year in Europe alone. However, research investment is comparatively low, despite the fact that it has the potential to substantially reduce this cost. In the UK, mental health problems were estimated to account for 22.8% of the disease burden, 7% more than cancer. However, cancer research receives 4.5 times more funding. The report concluded that substantially more mental health research and research funding should be allocated to the areas of priority they identified, co-morbidity being one of them.

For the abovementioned reasons, various co-morbid mental health disorders warrant further research, but the subject of this thesis is the specific relationship between Major Depressive Disorder (MDD) and Cannabis Use Disorder (CUD), because these disorders co-occur at a rate greater than chance (Degenhardt, Hall, Lynskey, Cofey, & Patton, 2012), are both relevant from a public health perspective (Torrens et al., 2015), and the mechanisms driving their co-morbidity have not been sufficiently explored.

1.1.1 Major Depressive Disorder

MDD is characterised by several emotional, cognitive and physiological symptoms, including depressed mood and/or the loss of pleasure or interest (American Psychiatric Association, 2013). Further symptoms include sleep disturbances, loss of energy, feelings of worthlessness or guilt, concentration problems, weight loss or gain, psychomotor agitation and suicidal thoughts. According to the DSM-5, five of these symptoms lasting for longer than two weeks – and including depressed mood and loss of pleasure – would qualify for a diagnosis of MDD (American Psychiatric Association, 2013).

MDD is a highly prevalent and debilitating condition for the individual. Lifetime prevalence estimates lie between 6.6% (Japan) and 21.0% (France; Kessler & Bromet, 2013). According to the latest Global Burden of Disease study in 2013, MDD

was the second leading cause of disability in the world (Ferrari et al., 2013), and the World Health Organisation is now citing it as the leading one (WHO, 2017). The disorder is associated with a variety of adverse consequences for the individual, among which are marital dissatisfaction, financial and educational difficulties (Kessler & Bromet, 2013), as well as impairments in cognitive functioning (Snyder, 2013).

The disorder is estimated to be a substantial economic burden on society (Greenberg, Fournier, Sisitsky, Pike, & Kessler, 2015). Among mental health conditions, the financial burden of mood disorders (including MDD) is highest, costing European countries €113.4 billion and making up almost a quarter of the costs of mental health conditions (Olesen, Gustavsson, Svensson, Wittchen, & Jönsson, 2012).

1.1.2 Cannabis use and Cannabis Use Disorder

Cannabis is the most commonly used illicit drug (Agrawal & Lynskey, 2014). In a survey of 15 to 16 year old American students, cannabis use even exceeded the prevalence of tobacco use (EMCDDA, 2017). The most recent EMCDDA drug report estimates that 26.3% of 15 to 65 year olds living in the European Union have used cannabis in their lifetime, 7% in the past year (EMCDDA, 2017). Clarity about its physical and mental health consequences is of particular relevance due to the ongoing debate around its legalisation (EMCDDA, 2017). Heavy cannabis use is linked to an increased likelihood of road accidents, lower educational attainment, financial difficulties, relationship difficulties, physical health complications (e.g. respiratory and cardiovascular problems), and various mental health disorders (Cerdá et al., 2016; Hall, 2015; Hall & Degenhardt, 2009). In Europe, the number of first-time treatment seekers for cannabis problems has almost doubled in the past decade (EMCDDA, 2017).

Some cannabis users will develop clinical levels of drug taking and drug seeking behaviour. In the most recent version of the DSM (DSM-5), clinical levels of cannabis involvement are defined as Cannabis Use Disorder, which is diagnosed when at least two symptoms of drug abuse, dependence or craving (see Table 1) occur

within the same 12-month period (American Psychiatric Association, 2013). In the previous version of the DSM (DSM-IV; American Psychiatric Association, 2000), cannabis dependence and abuse were diagnosed separately, but involved almost identical criteria. The estimated number of individuals who transition from use to use disorder, abuse or dependence varies by cohort. A recent analysis of the 2007 National Survey of Mental Health and Wellbeing in Australia found that 22.8% of cannabis users developed abuse, and 9.8% developed dependence symptoms (Degenhardt et al., 2018). Worldwide an estimated 13.1 million individuals are cannabis dependent, which contributes to approximately 2.1 million years of healthy life lost due to disability (Degenhardt, Whiteford, & Hall, 2014).

Table 1. Symptoms of DSM-IV and DSM-5 definitions of clinical cannabis involvement. Adapted from Hasin et al. (2013).

Symptoms	DSM-IV Abuse	DSM-IV Dependence	DSM-5 Substance Use Disorder
Hazardous Use	X	-	X
Social/interpersonal problems related to use	X	-	X
Neglected major roles to use	X	-	X
Legal problems	X	-	-
Withdrawal	-	X	X
Tolerance	-	X	X
Used larger amounts/longer	-	X	X
Repeated attempts to quit/control use	-	X	X
Much time spent using	-	X	X
Physical/psychological problems related to use	-	X	X
Activities given up to use	-	X	X
Craving	-	-	X

1.1.3 Co-morbidity between Major Depressive Disorder and Cannabis Use Disorder

The relationship between MDD and CUD is poorly understood. Cross-sectional studies of general and clinical populations consistently show that clinical and sub-clinical symptoms of CUD and MDD co-occur at a rate greater than chance (see Degenhardt et al. 2012 for a review). This co-morbidity is highly relevant from a

public health perspective, because of the aforementioned burden of both disorders individually, as well as the added burden of co-morbidity. However, research into aetiological mechanisms, including longitudinal and genetic studies, has so far not succeeded in clarifying the underlying causes of the relationship between MDD and CUD. Developing a greater understanding of it is important to help reduce the prevalence of both conditions, and associated risks, through the efficient prevention and treatment of co-morbid cases.

This chapter will provide an overview of existing research into the co-morbidity between CUD and MDD, introducing cross-sectional studies demonstrating significant co-morbidity, then moving on to outline possible mechanisms of co-occurrence. It will outline the limitations of the studies conducted so far and explain how the current thesis will attempt to address gaps in the literature.

1.2 Reviewing the literature: Definitions of cannabis involvement and depression

The subject of this thesis is psychiatric co-morbidity, as outlined above, so the phenotypes of interest are MDD and CUD. However, limiting the literature review to only CUD and MDD would severely restrict the number of informative studies that can be reviewed. To encompass the various definitions of related phenotypes, this introduction will frequently refer to evidence of the association between ‘cannabis involvement’ and ‘depression’, and not limit itself to CUD and MDD only.

The definitions of cannabis involvement and depression used in the literature are heterogeneous, so that non-clinical levels of cannabis use encompasses different definitions of frequency (e.g. ‘daily’ vs ‘heavy’) and clinical levels of cannabis involvement and depression are often not measured as CUD and MDD. Clinical levels of cannabis involvement are most frequently defined as cannabis abuse and/or dependence, since CUD was only introduced in the latest version of the DSM (American Psychiatric Association, 2013). Abuse and dependence criteria are still core components of a CUD diagnosis and therefore remain relevant. Studies examining clinical levels of depression may include dysthymia or bipolar disorders

alongside MDD as a ‘mood disorder’ category, use a subset of MDD symptoms, and measure depression using different instruments.

Overall, studies which have not specifically examined CUD and MDD, but have looked at related phenotypes (e.g. frequent cannabis use or dysthymia) can still provide information on the likely co-morbidity between CUD and MDD, because they are thought to be sufficiently related to the definitions used in this thesis. To capture all relevant evidence, a comprehensive literature review on the relationship between depression and cannabis involvement was conducted in PubMed, which included the following search terms: ‘cannabis’, ‘marijuana’, ‘marihuana’, ‘tetrahydrocan*’, ‘CBD’, ‘THC’ or ‘cannabinoid*’, and the words ‘MDD’, ‘major depress*’, ‘depress*’, ‘anhedoni*’, ‘dysthymi*’, ‘low mood*’ or ‘mood disorder*’. Since informing on the mechanisms of co-occurrence rests on an exploration of causality, the target of this review were primarily twin and longitudinal studies, so the search also included the terms ‘longitudinal’ or ‘twin’.

Because there were few twin studies, and a recent meta-analysis (Lev-Ran et al., 2014) had been published on the longitudinal relationship between cannabis involvement and depression at the time of the first literature search in 2014, a systematic literature review has not been conducted. However, the same search terms have been used to update the literature throughout the course of this thesis (between 2014 and 2018).

Therefore the following discussion of cross-sectional studies on the extent of the co-morbidity between MDD and CUD is followed by a discussion of possible mechanisms of co-occurrence, which contains a comprehensive and current account of twin and longitudinal studies on the causal relationship between cannabis involvement and depression.

1.3 Significant co-morbidity? Evidence from cross-sectional studies

As mentioned above, the co-morbidity between cannabis involvement and depression has been consistently observed (see Degenhardt et al. 2012 for a

review). For instance, an epidemiological study of 43,093 US citizens showed that individuals with mood disorders (MDD, dysthymia, mania, hypomania) had 3.9 (95% CI = 2.8 – 5.3) times higher odds of meeting the criteria for lifetime cannabis abuse or dependence than individuals without mood disorders (Martins & Gorelick, 2011). An epidemiological survey of 25,113 Canadian citizens reported that rates of past-year cannabis dependence among individuals who met past-year MDD criteria were over 7.25 times higher than those of individuals who did not (Patten et al., 2015). Rates of past-year cannabis abuse were almost 3.6 times higher (Patten et al., 2015). Similar results have been found in clinical samples. For example, a recent study based on the Norwegian patient registry including 2,659,966 individuals reported that levels of ICD-10 depressive illness were almost 3.9 times higher among individuals with CUD (12.85%), compared to the general population (3.3%, Nesvåg et al. 2015).

These cross-sectional studies, in conjunction with several others (C.-Y. Chen, Wagner, & Anthony, 2002; Cougle, Hakes, Macatee, Chavarria, & Zvolensky, 2015; R. R. S. Mathews, Hall, & Gartner, 2011) present convincing evidence that clinical levels of depression and cannabis involvement co-occur significantly more frequently than expected by chance, which raises questions about the possible mechanisms that may underlie this association.

1.4 Mechanisms underlying co-morbidity

There are multiple mechanisms which could explain the observed co-morbidity. Firstly, some co-morbidity in a sample is expected due to chance: it is the product of the prevalence of each disorder separately. Secondly, sampling bias is likely to lead to inflated estimates of co-morbidity in clinical samples. Co-morbid individuals are likely to have more severe symptoms than individuals with pure disorders and would therefore be more likely to enter a clinic (Berkson, 1946). Additionally, clinics in which practitioners have a special interest in particular disorders may have an increased rate of referrals for clusters of these disorders (Caron & Rutter, 1991). Further processes by which such ‘artefactual co-morbidity’ may occur are reviewed in Caron and Rutter (1991). Overall, the highest likelihood of observing true co-

morbidity is in general population samples, which is what most evidence reviewed here focuses on.

Most research on the mechanisms underlying true co-morbidity has examined overlapping risk factors and causal processes. Research into overlapping risk factors aims to find the *source* (i.e. biological versus environmental), of the overlap between risk factors influencing two disorders (e.g. MDD and CUD). Studies often remain agnostic about the *process* by which this source gives rise to co-morbidity. One possibility is that overlapping risk factors simultaneously increase the likelihood of developing both disorders. However, these risk factors may also be involved in other processes, including, but not limited to, causality. As an example, the endocannabinoid system is both involved in mood regulation and is the primary site of action for cannabinoids (see Section 1.4.1.1). Any malfunctioning of the endocannabinoid system, perhaps due to genetic factors, may increase the risk of CUD and MDD simultaneously. However, the endocannabinoid system may also provide a plausible site through which mood problems affect cravings for cannabinoids, or cannabis consumption alters mood (i.e. causality). Still, causality is not the only mechanism via which co-morbidity can arise and there are several different categories of causal processes (Caron & Rutter, 1991; Klein & Riso, 1993; Neale & Kendler, 1995). An overview of these alternative mechanisms can be found in Sections 1.4.3 (Chapter 1), 3.8.2 (Chapter 3) and Chapter 6.

The following sections will outline potential overlapping risk factors, summarise research into causal mechanisms and give an overview of other potential explanations of co-morbidity.

1.4.1 Overlapping risk factors

Overlapping risk factors may be biological or environmental, and individuals who are exposed to or inherit these risk factors may simultaneously be at increased risk for MDD and CUD. Alternatively, overlapping risk factors may be a site of action for causal or other non-causal aetiological processes.

1.4.1.1 Biological overlap between CUD and MDD - the endocannabinoid system

Research from various fields suggests that the main site of biological – and potentially genetic – overlap between cannabis-related problems and symptoms of depression is likely to be the endocannabinoid system. It is the primary site of action for cannabis and contains two types of G protein-coupled receptors: CB_1 and CB_2 (Mechoulam & Parker, 2013). These receptors are activated by cannabinoids endogenous to the human organism, mainly anandamide and 2-arachidonoyl glycerol (2-AG), as well as synthetic and plant-derived cannabinoids. Delta-9-tetrahydrocannabinol (THC), the principal psychoactive component of cannabis binds to both endocannabinoid receptors. In the central nervous system, endocannabinoids act on CB_1 receptors to modulate the release of various neurotransmitters and other neuromodulators (Gorzalka, Hill, & Hillard, 2008). The binding of THC to CB_1 receptors is thought to be responsible for the psychological effects of cannabis, but the endocannabinoid system can also be linked to MDD (Agrawal et al., 2012; Gorzalka et al., 2008; Hill et al., 2008; Mechoulam & Parker, 2013). The pathophysiology of MDD seems to involve complex interactions between monoamine signalling, neurogenesis, inflammation and the human stress response. Endocannabinoids interact with or modulate each of these processes, providing a possible biological and genetic link between CUD and MDD. The main site of these interactions is likely to involve the hypothalamic–pituitary–adrenal axis and limbic brain regions in general, which include, but are not limited to the hippocampus, amygdala, and nucleus accumbens (Patel & Hillard, 2009).

1.4.1.2 Environmental overlap between CUD and MDD

There are also several plausible environmental factors which could predispose an individual to both disorders simultaneously. Individuals with both MDD and cannabis-related problems are likely to have experienced higher levels of childhood adversity (Fergusson & Horwood, 1997; Heim & Nemeroff, 2001; Nanni, Uher, & Danese, 2012; Nelson et al., 2006), which is likely to occur before the onset of either disorder and predispose a person to both. Additionally, cannabis users (Fergusson & Horwood, 1997; Georgiades & Boyle, 2007) and individuals showing symptoms of MDD (E. Chen & Miller, 2013; Gilman, Kawachi, Fitzmaurice, & Buka, 2002;

Sadowski, Ugarte, Kolvin, Kaplan, & Barnes, 1999) are more likely to come from a disadvantaged social background. Risk factors associated with both are also likely to include problematic peer relationships, educational attainment, traumatic life events, and marital status (Degenhardt et al., 2012; Manrique-Garcia, Zammit, Dalman, Hemmingsson, & Allebeck, 2012).

Table 2 summarises the covariates that show a significant association with depression, cannabis involvement or both, in studies which examine the relationship between the two phenotypes. The summary focuses on longitudinal – rather than cross-sectional – studies, since they better differentiate between covariates that are likely to precede and those that are likely to follow the onset of depression and cannabis involvement. Covariates which precede their onset are of primary interest because they may provide insight on mechanisms leading to co-morbidity.

Table 2. Covariates of depression, cannabis involvement and both, identified from longitudinal studies.

Study	Both	Depression	Cannabis involvement
Fergusson and Horwood (1997)	As a block[1]: gender, changes of parents (0–15 years), childhood sexual abuse, IQ age 8, mood disorder age 14	Not reported	Family social background (mother aged <25 at birth of child, mother had no formal educational qualification, family of semi-skilled/unskilled SES), family functioning (> 2 changes of parents, parental conflict, parental history of offending, parental history of alcoholism/alcohol problems, parental illicit drug use, exposed to childhood sexual abuse), individual factors (conduct problems, self-esteem, novelty seeking), parent/peer relationships (parental attachment, deviant peers), adjustment prior to age 16 (anxiety disorder, alcohol abuse, other illicit substance use, daily smoking, history of property or violent offending)
Bovasso (2001)	Not reported	None of the measured variables	Not reported
Brook et al. (2002)	None	Family income, parental education level	None
van Laar et al. (2007)	Age, gender, neurotic personality, parental psychiatric history, traumatic events in childhood	Not reported	Not reported
Georgiades and Boyle (2007)	Not reported	Not reported	Age, family SES, single parent home, family dysfunction, internalising disorders, externalising disorders, grade failure
Pedersen (2008)	As a block[1]: age, gender, parental educational level, parents unemployed or receiving social welfare benefits, parental divorce, parental smoking and alcohol problems, parental support and monitoring measured, early pubertal maturation, school marks, conduct problems and daily smoking, alcohol intoxication, alcohol problems, impulsivity	Father unemployment, parental divorce, both parents smoking, poor parental support, poor parental monitoring, early pubertal maturation, low school marks, conduct problems, daily smoking, alcohol problems, no full secondary education, living on social security	None reported

Table continues on next page

Study	Both	Depression	Cannabis involvement
Harder et al. (2008)	Not reported	Not reported	Gender, race, daily tobacco smoking, alcohol abuse or dependence, other illegal drug use, concentration problems, behaviour problems
Marmorstein and Iacono (2011)	Educational/occupational failure (not graduating from high school/earning GED by age 20, a period of unemployment for at least 6 months when wanting to work), crime (trouble with police, aside from traffic stops)	Same as 'both'	Same as 'both'
Brook et al. (2011)	Not reported	Not reported	Gender, age, violence towards others, psychological symptoms, peer deviance, association with drug using peers, skipped work, violence towards subject in neighbourhood
Manrique-Garcia et al. (2012)	As a block[1]: prior personality disorders at conscription, IQ, disturbed behaviour in childhood, social adjustment, risky use of alcohol, smoking, early adulthood, socioeconomic position, use of other drugs, brought up in a city.	Personality disorders at conscription, IQ, smoking, disturbed behaviour in childhood	Not reported
Degenhardt et al. (2013)	Age, level of education, nationality, parental smoking status, tobacco use, symptoms of anxiety, maximum qualification achieved (secondary education, vocational qualification, degree), ever had a baby, currently partnered/married, receiving government welfare, in paid employment.	Not reported	Not reported
Pacek et al. (2013)	Not significantly diminished	Gender, age, marital status, ethnicity, years of education, other drug use disorders	Gender, age, marital status, ethnicity, years of education, income, family history of depression
Cerdá et al. (2013)	Ethnicity, family SES	Not reported	Not reported
Haug et al. (2014)	Not reported	Not reported	Means of subsistence, having siblings, belief in God and practicing religion, parental divorce before age 18, higher parental knowledge of peers and whereabouts at age 15, peer pressure, nicotine dependence, sensation seeking, anxiety

Table continues on next page

Study	Both	Depression	Cannabis involvement
Silins et al. (2014)	As a block[1]: all 53 covariates included, see reference for details	Alcohol use, father smoking status, anxiety symptoms	Gender, highest maternal education, other illicit drug use before age 17
Feingold et al. (2015)	Not significantly diminished	Not reported	Gender, household income, marital status, age
Cougle et al. (2015)	As a block[1]: age, income, marital status, gender, ethnicity, education, and psychiatric co-morbidity (anxiety disorder, personality disorder, bipolar disorder)	Not reported	Not reported
Gage et al. (2015)	Cigarette, alcohol or other illicit drug use	Not reported	Low maternal education, IQ at age 8, conduct disorder, cigarette smoking at age 16, illicit drug and alcohol use at age 16
Blanco et al. (2016)	Non-significant OR and aOR	Not reported	Childhood (family history of alcohol and illicit drug use disorders, disturbed family environment, childhood parental loss), early adolescence (low self-esteem), early-onset anxiety, social deviance), late adolescence (past year trauma, Axis I co-morbidity, Axis II co-morbidity), adulthood (divorce, history of alcohol, drug use disorder or nicotine dependence, social deviance), socio-demographics (age, gender)
Danielsson et al. (2016)	Age, sex, family tension, alcohol use, other illicit drug use	Not reported	Sex, age, education, place of upbringing, family tension, substance use (alcohol and other illicit drugs)

Note. 1. Further information on these studies, including cohort sizes and types of measures used are presented in Tables 5 and 6 for studies since 2013. Earlier studies are presented comprehensively in Lev-Ran et al. (2014).

2. Some longitudinal studies do not report any effects of single covariates and are therefore not listed here. Predictors were included if they were reported to a) significantly attenuate the relationship between depression and cannabis involvement after being included, b) have a significant association with either cannabis involvement or depression. Also, they were included if there was a non-significant association between depression and cannabis involvement when predictors were included immediately (only aOR reported).

[1]As a block = it is unclear which specific variable significantly attenuated association, but there was an attenuation after controlling for a group of variables

1.4.1.3 Testing environmental and biological overlap

Convincing evidence identifying significant biological and environmental risk factors for MDD and CUD comes from cross-sectional and longitudinal studies (see Section 1.4.2.5) which have controlled for potential covariates and observed a significant decrease in the strength of association between CUD and MDD, or related phenotypes. In a cross-sectional study, Degenhardt, Hall and Lynskey (2001) found that the odds ratios (ORs) between cannabis use, abuse and dependence and affective disorders (MDD, dysthymia, bipolar I and bipolar II) all become non-significant after other drug use (tobacco, alcohol and illicit drugs) had been controlled for. In a longitudinal study of 85,088 individuals across 17 countries, De Graaf et al. (2010) initially found a significant increase in the risk of developing symptoms of MDD after early-onset (before age 17) cannabis use, which became non-significant when childhood conduct problems were accounted for. In a longitudinal study conducted by Fergusson and Horwood (1997), it was found that childhood sexual abuse, gender, changes of parents by the age of 15, IQ and mood disorders at baseline significantly affect the association between cannabis use frequency and MDD in late adolescence (16 to 18 years). When controlling for these and other family and childhood factors, the association between cannabis use frequency and MDD became non-significant. A recent longitudinal study by Danielsson, Lundin, Agardh, Allebeck and Forsell (2016) confirmed the previously identified importance of family problems and other drug use. After controlling for family tensions, illicit drug and alcohol use, cannabis use at baseline no longer appeared to be significantly associated with MDD at follow-up.

All of the abovementioned factors seem to overlap between cannabis and depression-related phenotypes and have the potential to explain some of the co-morbidity between CUD and MDD. However, cross-sectional and longitudinal studies cannot answer whether their influence could occur via environmental or genetic pathways. For instance, other drug dependence is likely to have both environmental (e.g. smoking cigarettes) and genetic (e.g. innate predisposition toward nicotine dependence) components. Additionally, there may be an interaction between environmental and genetic components, since nicotine consumption (environmental)

may affect the read-out of certain genes via epigenetic mechanisms (Levine et al., 2011), which in turn affect the development of CUD and MDD.

Determining whether the covariance between CUD and MDD is primarily explained by genetic or environmental factors would require a genetically informative study design, such as twin studies. It is an important empirical question, since knowing the aetiological mechanisms of the co-morbidity can guide further research, as well as prevention and treatment of co-morbid cases. Should the correlated risk factors be primarily genetic, family history methods (Milne et al., 2009) and genetic risk scores based on molecular genetic studies might be the best way to screen for at-risk individuals for preventive interventions before they develop a co-morbidity. If overlapping risk factors were primarily environmental, financial resources should be allocated toward environmental interventions. Additionally, screening for at-risk individuals for potential preventative measures should be primarily conducted in high-risk environments.

1.4.1.4 Twin studies as an avenue to test for genetic versus environmental overlapping risk factors for CUD and MDD

Twin studies are one of the most frequently used methods to decompose individual differences – or the variance – within a trait into genetic and environmental contributions. In contrast to a population of unrelated individuals, twin data contain genetic information: MZ twins share 100% of their DNA, while DZ twins share around 50% of their segregating genes. Segregating genes are those which contribute to individual differences in the population. As there are several reasons to assume that environmental influences do not significantly differ for either twin type (discussed in Chapter 3, Section 3.7.1; Rijsdijk & Sham, 2002), differences in the concordance of a trait between MZ and DZ twins raised together are most likely due to genetics. Using structural equation modelling (see Chapter 3, Section 3.5) it is then possible to capitalise on the differences in genetic resemblance between the twin types, to estimate how far traits, and the relationships between them, are influenced by genetic and environmental factors.

Twin studies so far have demonstrated considerable evidence that both MDD (e.g. Kendler, Gatz, Gardner, & Pedersen, 2006; Sullivan, Neale, & Kendler, 2000) and cannabis involvement (Lynskey et al., 2002; Verweij et al., 2010, 2013) are influenced by genetic factors (see Table 3). Significant genetic influences have been found for cannabis use initiation (Verweij et al., 2010), symptoms of abuse or dependence (Lynskey et al., 2002; Verweij et al., 2010), as well as cannabis withdrawal (Verweij et al., 2013). Since abuse and dependence are core components of CUD, the proportion of CUD variance explained by genetic factors is also likely to be around 50%. Environmental factors, which contribute to similarities (shared environmental factors; C) and differences (unique environmental factors; E) between individuals, are also likely to influence CUD. MDD is likely to be influenced primarily by environmental factors that contribute to differences between people (E).

Table 3. Estimates of genetic and environmental influences on MDD and cannabis involvement.

Phenotype	Study	Gender	A (95% CI)	C (95% CI)	E (95% CI)
MDD	Sullivan et al., 2000 (meta-analysis)	All	37% (31 - 42%)	0% (0 - 5%)	63% (58 - 67%)
	Kendler et al., 2006	Female	42% (36 - 47%)	Dropped from model (non-significant)	58% (53 - 64%)
		Male	29% (19 - 38%)	Dropped from model (non-significant)	71% (62 - 81 %)
Cannabis initiation	Verweij et al., 2010 (meta-analysis)	Female	40% (30 - 49%)	39% (29 - 49%)	21% (16 - 27%)
		Male	48% (38 - 58%)	25% (11 - 39%)	27% (22 - 32%)
Cannabis abuse/ dependence	Lynskey et al., 2002	All	45% (15 - 72%)	20% (0 - 44%)	35% (26 - 46%)
	Verweij et al., 2010 (meta-analysis)	Female	59% (44 - 73%)	15% (1 - 30%)	26% (23 - 30%)
		Male	51% (38 - 65%)	20% (11 - 28%)	29% (22 - 35%)

Note. A = proportion of phenotypic variance explained by genetic (heritable) factors. C = proportion of phenotypic variance explained by environmental factors shared between twins. E = proportion of phenotypic variance explained by non-shared environmental factors and measurement error. CI = Confidence Interval.

However, few twin studies so far have examined to what extent the *covariance* between MDD and CUD may be influenced by genetic and environmental factors. In a sample of Australian twins Lynskey et al. (2004) reported a significant genetic correlation between lifetime MDD and cannabis dependence (men: r = 0.44, 95% CI = 0.17 - 1.00; women: r = 0.69, 95% CI = 0.30 - 1.00). Part of that common genetic liability between MDD and cannabis dependence may be explained by genetic influences linked to Antisocial Personality Disorder (APD). Fu et al. (2002) conducted twin model analyses including multiple variables. Within these models, the genetic factors influencing APD were entered into the model first, accounting for a large portion of variance in cannabis dependence. This substantially reduced the amount of variance that was left to be accounted for by genetic factors related to MDD: while the genetic influences of MDD and cannabis dependence overlapped, 62% of this genetic correlation could be explained by APD. Once APD was taken into account in the genetic model, the genetic correlation between cannabis dependence and MDD became no longer significant.

The author is currently not aware of any other twin study reporting on genetic correlations between MDD and clinical cannabis involvement. Conducting such research would be important, since neither twin study was primarily investigating the bivariate genetic and environmental correlations between cannabis dependence and MDD. Lynskey et al. (2004) primarily conducted a discordant twin study and did not report on environmental correlations between the phenotypes. Fu et al. (2002) focused on the multivariate associations between APD, cannabis dependence, MDD and alcohol dependence. Additionally, twin study estimates are known to depend on geographical locations and time periods (Neale & Maes, 2004) and need to be replicated across these factors in order to be generalisable.

1.4.1.5 Conclusions and limitations – overlapping risk factors

Overall, there is evidence suggesting that cannabis involvement and depression may overlap neurobiologically and share environmental influences. The importance of these overlapping risk factors is unclear to date. Twin studies are capable of estimating the degree of genetic and environmental overlap between two traits without measuring specific genetic, biological or environmental factors. However,

evidence from twin studies on the bivariate relationship between CUD and MDD is lacking. While twin studies have identified that MDD and CUD are both likely influenced by genetic factors, further bivariate twin studies would be necessary to examine to what extent the *covariance* between MDD and CUD is influenced by genetic and environmental factors.

1.4.1.6 Limitations to be addressed in the current thesis

Twin studies that have been conducted so far have examined cannabis dependence and MDD, leaving uncertainty about the genetic and environmental relationship between MDD and clinical cannabis involvement when abuse criteria are included. Additionally, no twin study so far has primarily investigated the bivariate genetic and environmental correlation between clinical cannabis involvement and MDD. Therefore, a bivariate twin study of MDD and CUD will be reported on in Chapter 5.

1.4.2 Causality

Given the strong evidence of a significant co-morbidity and several possible sources of overlap between CUD and MDD, causality is of primary interest as a possible process of co-morbidity. A causal link would be of interest for policy, treatment and prevention. Cannabis legalisation is currently highly debated in various countries (Pacula & Smart, 2017). If any cannabis-related phenotype, such as CUD should causally influence a disorder as prevalent and disabling as MDD, this may influence whether and how cannabis would be legalised. From a treatment and prevention perspective, this finding could also clarify how to approach co-morbid cases of CUD and MDD. If CUD causes MDD, then it may be most efficacious to focus on treating the drug problem first. Conversely, finding that MDD causes CUD would suggest that co-morbid CUD and MDD cases may be treated best by addressing MDD primarily.

This causality may be uni- or bi-directional and involve biological or environmental mechanisms driving any causal association (Agrawal & Lynskey, 2014). High levels of cannabis use in CUD may alter the neural mechanisms the psychoactive components bind to, and these alterations may increase the risk for or induce MDD. Neurochemical changes occurring during the course of a major depressive episode

may make an individual more susceptible to use high levels of cannabis for mood elevation, which would be considered 'self-medication'. High levels of cannabis use may also lead to changes in the individual's environment, such as relationship and financial difficulties (Cerdá et al., 2016), which in turn act as a causal factors for MDD. Conversely, MDD may cause individuals to withdraw from social activities, leading to feelings of loneliness or social exclusion, against which cannabis may be used as a buffer (Deckman, DeWall, Way, Gilman, & Richman, 2014).

1.4.2.1 Causality: Evidence from twin studies

Twin studies are a powerful tool to examine whether causality is a possible explanation for the co-morbidity between two phenotypes. In a discordant twin design (see Table 4), the prevalence of an outcome (e.g. MDD) is compared in MZ twin pairs where only one twin pair member has been affected by a predictor (e.g. CUD). If the MZ twin who is affected by the predictor shows a significantly elevated rate of outcome (i.e. OR > 1), the association between MDD and CUD may be causal. This is because MZ twins who were raised together overlap 100% in their genetic and shared environmental influences. Comparing outcome rates in MZ twins discordant for a predictor therefore controls for all genetic and shared environmental influences. If the OR between CUD and MDD is elevated despite controlling for these influences, two conclusions are possible. Firstly, there may be a causal relationship between CUD and MDD. Secondly, a non-causal overlap in unique environmental influences, which do differ between MZ twins, may simultaneously increase the prevalence of CUD and MDD. Since discordant twin designs do not differentiate between these two possibilities, they do not prove causality.

However, if the OR is not significantly elevated in MZ twins discordant for CUD (i.e. OR = 1), this would mean that, after controlling for all genetic and shared environmental factors, there has been no causal association detected between MDD and CUD. While discordant twin analyses in MZ twins cannot prove causality, they can disprove it.

Unrelated individuals do not share genetic or any environmental influences. Consequently, any significant relationship observed in a non-MZ sample, e.g. a

sample of unrelated individuals usually enrolled in cross-sectional or longitudinal studies, could have occurred due to a non-causal correlation in genetic or shared environmental factors between MDD and CUD.

Discordant twin studies are an efficient way to examine whether causality is a likely explanation for the co-morbidity between CUD and MDD which was observed in cross-sectional studies of unrelated individuals, since both MDD and CUD have shown to be significantly influenced by genetic factors and twin studies have the critical advantage of controlling for all genetic and a large portion of environmental influences without having to measure them.

Table 4. Discordant twin design explained, based on rates of CUD and MDD in MZ twins, and pairs of unrelated individuals. Individuals are discordant for both predictor (e.g. CUD) and outcome (e.g. MDD) measures.

Relatedness	Person	CUD	MDD (OR > 1)	Reason for OR > 1	MDD (OR = 1 in MZ)
MZ	**A**	**Yes**	**Yes[1]**	**Overlapping unique environments and causality[1]**	**No[3]**
	B	**No**	**No[1]**		**Yes[3]**
Unrelated	A	Yes	Yes[2]	Overlapping genes, shared or unique environments and causality[2]	Yes
	B	No	No[2]		No

Note.
[1]MZ twins share 100% of their genetic and shared environmental influences, but 0% of their unique environmental influences. Therefore, a significantly increased likelihood of meeting MDD criteria in co-twins *with* CUD (OR > 1) must be due to two processes which are not controlled for in pairs of MZ twins: overlapping unique environmental influences or causality.
[2]Unrelated individuals share no genetic or environmental influences. Any sig. association between CUD and MDD could be explained by an overlap in all influences, or causality.
[3]If an OR > 1 has been found in unrelated individuals, causality is a possible explanation for co-morbidity. However, if MZ twins with CUD are *not* more likely to meet criteria for MDD than their co-twin without CUD (OR = 1), the MZ result disproves any causality suggested in populations of unrelated individuals. The difference in OR between unrelated individuals and MZ twins is likely due to overlapping genetic and/or shared environmental influences, which are controlled for in MZ twin comparisons.

1.4.2.2 Cannabis involvement causes depression

Discordant twin studies have produced mixed evidence with respect to potential causal mechanisms between cannabis involvement and depression. One Australian twin study examined whether the prevalence of lifetime cannabis dependence in one MZ twin increased their risk for subsequent MDD, compared to their co-twin who did not have a history of cannabis dependence (Lynskey et al., 2004). Conditional

logistic regressions were used to test whether cannabis dependence was associated with an increase in MDD in MZ twin pairs. The levels of MDD were not significantly increased in MZ twins with cannabis dependence, relative to their non-cannabis dependent co-twin (aOR = 1.16, 95% CI = 0.64 – 2.17). However, cannabis dependence was associated with elevated rates of MDD in DZ twin pairs discordant for cannabis dependence (aOR = 3.40, 95% CI =1.91 – 6.05). The absence of a significant association in MZ twin pairs suggests that cannabis dependence is unlikely to be a unique causal factor for MDD but that shared genetic influences, which are not fully controlled for in DZ twins, explain some component of the observed co-morbidity.

However, the absence of an association in MZ twins may also have been due to the relatively small sample size (277 twins discordant for cannabis dependence). A subsequent analysis (Agrawal et al., 2017) with a substantially larger sample size, including this and two additional Australian twin cohorts, found that MZ twins who had used cannabis over 100 times had significantly increased odds of MDD compared to their co-twins who were non-users or lighter users (aOR = 1.72, 95% CI = 1.05 – 2.82).

1.4.2.3 Depression causes cannabis involvement

Lynskey et al. (2004) also examined whether twins meeting the criteria for MDD before age 17 were more likely to subsequently report cannabis dependence than their co-twins without MDD. Similar to their results on the reverse association, there was a significantly elevated prevalence of cannabis dependence among twins with MDD in DZ (OR = 9.50, 95% CI = 2.21 – 40.78), but not MZ twins (OR = 1.38, 95% CI = 0.55 – 3.42). In other words, MDD was not shown to be causally associated with later cannabis dependence. However, Lin et al. (1996) found that the prevalence of cannabis abuse/dependence was significantly higher in MZ co-twins who met the criteria for MDD (aOR = 2.3; 95% CI = 1.1 – 4.7), relative to their co-twin who did not meet the criteria for MDD. To the best of the author's knowledge, no other discordant twin studies have been published to date.

1.4.2.4 Conclusions and limitation to be addressed in this thesis

Discordant twin studies so far have produced mixed evidence with respect to a possible causal mechanism between depression and cannabis involvement. Causal mechanisms cannot be discounted, particularly between CUD and MDD. To date, Lin et al (1996) is the only co-twin control study incorporating both cannabis dependence and abuse in their definition in a discordant twin analyses. However, their assessment was based on DSM-III-R phenotypes, their sample was relatively small (234 twin pairs discordant for MDD) and only included male veterans. A replication of their results extending to a general population sample of twins, males and females, and utilising more recent definitions of both CUD and MDD would be beneficial. This thesis will therefore explore the relationship between CUD and MDD using a discordant twin design (Chapter 5).

1.4.2.5 Causality: Evidence from longitudinal studies

1.1.1.1.1 Cannabis involvement causes depression

Causality has also been tested in longitudinal studies. Although longitudinal studies are more limited than twin studies in their ability to control for genetic and environmental factors, or test for biological and environmental processes of causality, they provide a strong test of an overall causal relationship. They have been conducted in both directions of effect, on a variety of populations, and have used different definitions of cannabis involvement and depression.

Longitudinal studies evaluating whether various forms of cannabis use, abuse or dependence are causally linked to symptoms or clinical diagnoses of depression have produced mixed results. Among studies which have controlled for symptoms or diagnoses of depression at baseline, the majority have found no significant link (Blanco et al., 2016; D. W. Brook, Brook, Zhang, Cohen, & Whiteman, 2002; J. S. Brook, Rosen, & Brook, 2001; Cougle et al., 2015; Danielsson et al., 2016; Degenhardt et al., 2013; Feingold, Weiser, Rehm, & Lev-Ran, 2015; Fergusson & Horwood, 1997; Gage et al., 2015; Georgiades & Boyle, 2007; Harder, Stuart, & Anthony, 2008; Harder, Morral, & Arkes, 2006; Manrique-Garcia et al., 2012; Paton,

Kessler, & Kandel, 1977; Pedersen, 2008; Silins et al., 2014), while some have (Bovasso, 2001; J. S. Brook, Zhang, & Brook, 2011; Marmorstein & Iacono, 2011; Pacek, Martins, & Crum, 2013; Silins et al., 2014; Van Laar, Van Dorsselaer, Monshouwer, & De Graaf, 2007).

A recent meta-analysis (Lev-Ran et al., 2014) of all but seven (Blanco et al., 2016; Cougle et al., 2015; Danielsson et al., 2016; Feingold et al., 2015; Gage et al., 2015; Pacek et al., 2013; Silins et al., 2014) of the abovementioned studies, has concluded that overall, there is a modest increased risk of developing depression following cannabis use (OR = 1.17, 95% CI = 1.05–1.30). This relationship is stronger for heavy cannabis use (OR = 1.62, 95% CI = 1.21–2.10). Although the relationship is statistically significant, it may be questionable whether these results indicate a true causal relationship between cannabis and depression. There was heterogeneity in estimates from individual findings (ORs range from 0.68 (95% CI = 0.20–2.33) to 4.00 (95% CI = 1.23–12.99), and only four out of 18 reported findings were statistically significant, which is a cause for concern. Ioannidis (2005) has argued that meta-analyses with a ratio of significant to non-significant relationships larger than 1:3 are not likely to have found true relationships.

Among several longitudinal studies which have been published since the meta-analysis (see Table 5), the author is only aware of two studies finding a significant adjusted association (Pacek et al., 2013; Silins et al., 2014), the strongest being between clinical levels of cannabis use (CUD) and depression (MDD; Pacek et al., 2013). This is compatible with findings from studies included in the meta-analysis (Lev-Ran et al., 2014): the strongest associations have been found in studies which examined the relationship between CUD and MDD (Marmorstein & Iacono, 2011) or depressive symptoms (Bovasso, 2001).

Overall, there is some evidence to support longitudinal causal associations between baseline clinical levels of cannabis involvement and later clinical levels of depressive symptoms, but it is likely that a significant association only exists among heavy users, who may also have clinical symptoms of CUD.

Table 5. Summary of longitudinal studies investigating the onset of depression (any) following cannabis use (any) since Lev-Ran et al. (2014).

Source	Country	Sample size	Follow-up period	Age of cannabis measurement (years)	Cannabis measure	Age of depression measurement (years)	Type of depression assessment	Depressive scale	Result (95% CI)
Danielsson et al. (2016)	Sweden	8598; 1275 cannabis users	3 years	20–64	Lifetime cannabis use	20–64	Current MDD symptoms	MDI (DSM-IV, ICD-9)	RR = 1.22 (1.06 – 1.42); aRR = 0.99 (0.82 – 1.17)
Blanco et al. (2016)	USA (NESARC)	34653; 1279 cannabis users	3 years	18–65+	Past year cannabis use: 'no use', 'some but less than once/month', '1 or more uses/month'	18–65+	12-month prevalence of MDD	AUDATIS-IV (DSM-IV-TR)	MDD prevalence at time 2: OR = 1.1 (0.9 – 1.2); aOR = 0.8 (0.7 – 1.0) MDD incidence at time 2: OR = 0.9 (0.7 – 1.1); aOR = 0.8 (0.7 – 1.0)
Cougle et al. (2015)	USA (NESARC)	34653; 671 weekly cannabis users	3 years	18–65+	Weekly cannabis use	18–65+	12-month prevalence of Depressive disorder (MDD + dysthymia)	AUDATIS-IV (DSM-IV-TR)	aOR = 1.07 (0.84 – 1.36)
Gage et al. (2015)	UK (ALSPAC)	4561 (1791 analysis sample)	2 years	16	Frequency of use '0 times', '1–20 times', '21–60 times', 'more than 60 times'	18	Unipolar depression ('mild', 'moderate', 'severe')	CIS-R (ICD-10)	OR = 1.50 (1.26 – 1.80); aOR = 1.30 (0.98 – 1.72)
Feingold et al. (2015)	USA (NESARC)	34653; 843 cannabis users, 59 MDD and any cannabis use	3 years	18–65+	Past year cannabis use; frequency from 'every day' to 'once a year', frequency of cannabis on a day when cannabis was used	18–65+	12-month prevalence of MDD	AUDATIS-IV (DSM-IV-TR)	OR (< weekly) = 1.18 (0.77 – 1.82); OR (weekly to less than almost daily) = 0.83 (0.50 – 1.39); aOR (daily to almost daily) = 0.94 (0.41 – 2.14) aOR (any) = 1.06 (0.76 – 1.49); aOR (< weekly) = 1.04 (0.65 – 1.68); aOR (weekly to less than almost daily) = 0.67 (0.37 – 1.22); aOR (daily to almost daily = 0.58 (0.22 – 1.51)

Table continues on next page

Source	Country	Sample size	Follow-up period	Age of cannabis measurement (years)	Cannabis measure	Age of depression measurement (years)	Type of depression assessment	Depressive scale	Result
Silins et al. (2014)	Australia (CHDS, VAHCS, ATP combined)	3765 analysis sample	Varied	Before age 17	Frequency of cannabis use before age 17: 'never', 'less than monthly', 'monthly or more', 'weekly or more', 'daily'	17–25	Past week to past month moderate or severe depression	CIDI, CIS, DASS	OR (< monthly) = 1.12 (1.01 – 1.25); OR (> monthly) = 1.26 (1.02 – 1.56); OR (> weekly) = 1.42 (1.03 – 1.94); OR (daily) = 1.56 (1.04 – 2.42) aOR (< monthly) = 1.01 (0.85 – 1.28); aOR (> monthly) = 1.01 (0.72 – 1.42); aOR (> weekly) = 1.02 (0.61 – 1.69); aOR (daily) = 1.02 (0.52 – 2.01)
Pacek et al. (2013)	USA (NESARC)	34653; 395 with CUD	3 years	18–65+	CUD: AUDADIS-IV substance use/substance use disorders	18–65+	Current MDD	AUDATIS-IV (DSM-IV-TR)	OR = 2.02 (1.35 – 3.04); aOR = 1.78 (1.17 – 2.71)

Note. MDD = Major Depressive Disorder, CUD = cannabis use disorder, DSM = Diagnostic and Statistical Manual of Mental Disorders, NESARC = National Epidemiological Survey of Alcohol and Related Conditions, ALSPAC = Avon Longitudinal Study of Parents and Children, CHDS = Christchurch Health and Development Study, VAHCS = Victorian Adolescent Health Study, ATP = Adolescent Temperament Project, MDI = Major Depressive Inventory, AUDATIS-IV = Alcohol Use Disorders and Associated Disabilities Interview Schedule-DSM IV Version, CIDI = Composite International Diagnostic Interview, CIS(-R) = Clinical Interview Schedule (Revised), DASS = Depression Anxiety Stress Scale, (a)OR = (adjusted) Odds Ratio, (a)RR = (adjusted) Relative Risk, CI = Confidence Interval, RR = Relative risk.

1.4.2.6 Depression causes cannabis involvement

There have also been several longitudinal studies investigating whether various forms of depression cause cannabis involvement. These studies have shown mixed results. Degenhardt et al. (2012) summarised the findings from 11 studies, 10 of which did not find a significant relationship. To the author's knowledge, five additional relevant studies have been published within the last six years, which have controlled for depression at baseline (see Table 6). As is evident from the table, three out of the five recent studies have found significant links between depression and cannabis involvement. This is true for the initiation of any cannabis use within the past 12 months (Feingold et al., 2015; Haug, López Núñez, Becker, Gmel, & Schaub, 2014), as well as the development of CUD (Pacek et al., 2013). However, Cerdá et al. (2013) and Danielsson et al. (2016) did not find a significant relationship. Overall, the results remain unclear. A recent review of the evidence has argued that the longitudinal evidence of mood disorders (including MDD) causing cannabis involvement is currently lacking (Lev-Ran & Feingold, 2017).

Table 6. Summary of longitudinal studies investigating the onset of cannabis use (any) following depression (any) since Degenhardt et al. (2012).

Source	Country	Sample size	Follow-up period	Age of cannabis measurement (years)	Cannabis measure	Age of depression measurement (years)	Type of depression assessment	Depressive scale	Result (95% CI)
Danielsson et al. (2016)	Sweden (PART)	8598	3 years	20–64	Initiation of any past year cannabis use	20–64	Current MDD symptoms	MDI (DSM-IV, ICD-9)	RR = 1.62 (1.28 – 2.03), aRR = 0.91 (0.72 – 1.16)
Feingold et al. (2015)	USA (NESARC)	34653; 3320 who have MDD but no cannabis use	3 years	18–65+	Initiation of any past year cannabis use	18–65+	12-month prevalence of MDD	AUDATIS-IV (DSM-IV-TR)	OR = 2.19 (1.54 – 3.13); aOR = 1.72 (1.10 – 2.69
Haug et al. (2014)	Switzer-land (C-SURF)	2774 men (who have not used cannabis at baseline), 30 individuals with MDD and cannabis use	15 months	19–20	self-report Q: any 12-month use of cannabis (yes/no)	19–20	12- month prevalence of MDD	self-report Q: MDI	aOR = 1.03 (1.01 – 1.05)
Pacek et al. (2013)	USA (NESARC)	34653; 3320 who have MDD but no cannabis use	3 years	18–55+	AUDADIS-IV CUD	18–55+	12-month prevalence of MDD	AUDATIS-IV (DSM-IV-TR)	OR = 2.01 (1.09 – 3.68); aOR = 2.28 (1.28 – 4.05)
Cerdá et al. (2013)	USA (PYS)	499 boys, half selected to have high antisocial behaviour scores	10 years	9–19 (semi-annually)	SRA (before age 10), 16-item Substance Use Scale (after age 10); timing, quantity and frequency of marijuana use	9–19 annually	Depression symptoms	13-item RMFQ	aHR (recent depression) = 0.99 (0.95 – 1.04); aHR (cumulative depression) = 1.00 (0.95 – 1.05)

Note. MDI = Major Depression Inventory, MDD = Major Depressive Disorder, CUD = cannabis use disorder, CD = Cannabis Dependence, CA = Cannabis Abuse, DSM = Diagnostic and Statistical Manual of Mental Disorders, MDE = Major depressive episode, SRA = Self-Report Antisocial Behaviour Scale, RMFQ = Recent Mood and Feelings Questionnaire, AUDATIS-IV = Alcohol Use Disorders and Associated Disabilities Interview Schedule-DSM IV Version, (a)OR = (adjusted) Odds Ratio, (a)RR = (adjusted) Relative Risk, aHR = adjusted Hazard Ratio, CI = Confidence Interval.

1.4.2.7 Longitudinal studies of causality – conclusions and addressed limitations

Results from longitudinal studies are both mixed and difficult to interpret due to their limitations. Firstly, measures of cannabis involvement and depression have been used inconsistently. Measures of depression vary from several symptoms assessed on a self-administered scale to clinical interviews. Cannabis use varies between any lifetime use to clinical cannabis abuse/dependence. Pooling between studies becomes difficult in the face of this heterogeneity, since it is unclear to what extent the results are comparable (Lev-Ran et al., 2014). Secondly, differences between users and non-users do not seem to be sufficiently accounted for. For instance, Feingold et al. (2015) found that cannabis users and non-users significantly differed in age, gender, household income and marital status. An early study by Fergusson and Horwood (1997) also demonstrated differences on a large amount of factors, including childhood adversities, social disadvantages, contact with peers who engaged in substance use or delinquent behaviours, and psychological adjustment problems. Similarly to several other studies reviewed above (Danielsson et al., 2016; Gage et al., 2015; Silins et al., 2014), Fergusson and Horwood's (1997) study first showed a significant association between cannabis involvement and later depression, but it became non-significant following adjustment for covariates. This highlights the importance of controlling for a large variety of confounding factors, many of which seem to exert an influence on the association between cannabis and depression.

Furthermore, many longitudinal studies investigating the effect of cannabis involvement on later depression have examined low levels cannabis involvement, which may not be likely to lead to symptoms of depression. Given the relatively low toxicity of cannabis compared to alcohol and tobacco (Nutt, King, & Phillips, 2010), it is difficult to explain why having used cannabis a few times during one's lifetime would lead to depressive symptoms, via biological or environmental pathways. Longitudinal studies which have examined the relationship between CUD and MDD report the highest ORs (Bovasso, 2001; Marmorstein & Iacono, 2011; Pacek et al., 2013). However, interpreting these results is challenging, because these studies

have also controlled for a smaller number of covariates compared to others (e.g. Silins et al., 2014).

The inconsistent controlling for covariates may be explained by data collection for several of the cohorts mentioned above since they were not focused specifically on cannabis or MDD. For example, Pacek et al.'s (2013) data analysis was conducted on a dataset collected for the study of alcohol-related problems. Consequently, such cohorts would not be enriched for known covariates of cannabis involvement and thus unable to adequately control for them. Further epidemiological analyses of cohorts in which data collection was centred around cannabis involvement or depression and associated covariates could help to clarify the role such covariates are likely to play in the co-morbidity between MDD and CUD.

Overall, the strongest relationships between depression and cannabis-related phenotypes were observed between clinical levels of cannabis involvement and depression. It is plausible that there is a causal relationship between these; it is also plausible that these strong relationships are due to non-causal overlapping risk factors, which were not sufficiently controlled for. Consequently, this thesis examines the relationship between MDD and CUD, controlling for a multitude of potential covariates. The results are presented in Chapter 4.

1.4.3 Alternative models of co-morbidity

Given that longitudinal and discordant twin studies so far have not been able to clarify the source of co-morbidity between cannabis involvement and depression in general, and CUD and MDD in particular, it is crucial to highlight that causality is only one of many aetiological mechanisms that can lead to co-morbidity (Klein & Riso, 1993; Neale & Kendler, 1995). Twin studies are a powerful method to examine mechanisms other than overlapping risk factors. Previous twin studies have examined 13 different models of co-morbidity in different phenotypes, of which causal and correlated risk factor models are only a subset (Agrawal et al., 2007; Agrawal, Neale, Prescott, & Kendler, 2004; Agrawal, Silberg, Lynskey, Maes, & Eaves, 2010; Neale & Kendler, 1995; Rhee et al., 2004). The models and the questions they investigate are summarised in Table 7.

Table 7. Co-morbidity models according to the Neale and Kendler (1995) approach.

Model	Question
1. Alternate Forms	Alternate forms of the same disorder?
2. Three Independent Disorders	Co-morbid form is an independent disorder?
3. Random Multiformity	Abruptly increase symptoms of each other?
4. RM of MDD	MDD abruptly increases CUD symptoms?
5. RM of CUD	CUD abruptly increases MDD symptoms?
6. Extreme Multiformity	CUD and MDD abruptly increase symptoms of each other in extreme cases?
7. EM of MDD	MDD abruptly increases CUD symptoms in extreme cases?
8. EM of CUD	CUD abruptly increases MDD symptoms in extreme cases?
9. Correlated Liabilities	Risk factors are correlated?
10. Reciprocal Causation	CUD and MDD cause each other?
11. MDD causes CUD	MDD causes CUD?
12. CUD causes MDD	CUD causes MDD?
13. Chance	Co-morbid due to chance?

Using structural equation modelling, these models examine whether:

a) CUD and MDD may be *alternate forms* of one underlying distribution of risk factors. If this was the best-fitting model, CUD and MDD should be regarded as one disorder which can manifest as either cluster of symptoms.

b) CUD and MDD are pure forms of a third, *unrelated* disorder. This model assumes that there are three independent liabilities for CUD, MDD and co-morbid CUD with MDD.

c) The liabilities for CUD and MDD are *unrelated*. CUD discontinuously increases the risk of MDD symptoms and vice versa when thresholds are crossed. *Random multiformity* assumes one threshold and *extreme multiformity* assumes two thresholds. For instance, there may not be an increased risk of MDD due to CUD risk factors up until some threshold (e.g. a certain level of heavy cannabis use), after which the risk of MDD suddenly increases.

d) The liabilities, or risk distributions, are correlated. In other words, this model investigates overlapping risk factors – which have been mentioned above.

e) CUD and MDD are causally related. This test is possible using cross-sectional data because different causal directions lead to a different predicted result pattern in twins.

Family, ideally twin, data is necessary to investigate and compare all 13 models. Comparison of the different models is possible because different models assume different patterns of co-morbidity in twins. The best fitting model will create an expected pattern of co-morbidity which most closely matches the pattern of co-morbidity that was observed.

Since no previous study has examined all models for MDD and CUD, this thesis will aim to address this gap in the literature, and the results of this investigation will be presented in Chapter 6. A more thorough explanation of all models can be found in Neale and Kendler (1995), and will follow in Chapters 3 (Section 3.8.2) and 6.

1.5 Conclusion and thesis aims

Based on the evidence outlined in this chapter, it continues to be important to investigate the co-morbidity between MDD and CUD, as previous research, including cross-sectional, longitudinal and genetically informative studies, has produced mixed evidence with respect to the aetiological mechanisms underlying this relationship. The current thesis will analyse data from a sample of 3824 twins and their non-twin siblings, born between 1972 and 1979 in Australia. Epidemiological and twin analyses will aim to address the gaps in the literature identified in the review above.

Firstly, epidemiological analyses (Chapter 4) investigating the co-morbidity between CUD and MDD will be conducted on a cohort which is enriched for cannabis-related covariates, as this was the central aim of data collection. Consequently, the association between MDD and CUD will be well-controlled for potential covariates. These analyses will identify covariates which are likely to play a particularly important role in explaining the co-morbidity of the two disorders. Epidemiological

analyses on CUD and MDD are of particular value to the existing literature because few studies have used these definitions of cannabis involvement and depression.

Secondly, bivariate twin models (Chapter 5) will provide an estimate of the heritability of CUD and MDD and assess the impact of genetic and environmental factors on the covariance between the two disorders. These analyses will be helpful in guiding further studies and help understand whether individuals at risk for the co-morbidity, and thus who may benefit from early intervention, are best identified based on genetic or environmental risk factors.

Thirdly, discordant twin analyses in Chapter 5 will examine whether a causal relationship between CUD and MDD is a plausible explanation of their co-morbidity. If the association between MDD and CUD is found to be significant in MZ twins, causality can be explored in further twin model analyses. These analyses are also a valuable addition to the literature, since there are currently only three discordant twin studies on cannabis involvement and MDD, only one of which examined cannabis abuse/dependence.

Lastly, this thesis will examine 13 different models of co-morbidity, including but not limited to, causal models (Chapter 6). One of the limitations of cross-sectional and longitudinal studies is the lack of genetic information, which can be utilised to powerfully examine various aetiological links between two disorders. Examining the 13 co-morbidity models proposed by Neale and Kendler (1995) will be a valuable addition to the literature, since causality and overlapping risk factors so far cannot fully explain the co-morbidity between CUD and MDD, and a comprehensive test of alternative models has not been conducted previously.

2 Methods

The methodology for this thesis will be outlined in two chapters: the current chapter, on general methods and Chapter 3, on twin methods. The current general methods chapter will present a detailed overview of the sample, phenotype and covariate measures. Additionally, it will introduce both epidemiology and discordant twin-related statistical analysis. The statistical analysis section will then be expanded on in the subsequent twin methods chapter, which aims to present a thorough explanation of twin methods in general and the specific approaches used for this thesis.

2.1 The sample

The analyses presented are based on a sample of Australian twins from the 'Australian Twin Study of Cannabis and other Illicit Drug Misuse'. Data collection was centred around cannabis involvement and its correlates, which included measures on upbringing, other drug involvement, other mental health conditions, and several other factors (Lynskey et al., 2012).

Participants from this sample were recruited through the Australian Twin Registry (ATR), a volunteer registry maintained by the University of Melbourne (Hopper, Foley, White, & Pollaers, 2013), currently consisting of around 74,000 members. The main responsibilities of the ATR were and are the recruitment of twins into the registry, the maintenance of their contact details and baseline questionnaires, and the facilitation of research on twins (Hopper, Treloar, de Klerk, & Morley, 2006). Recruitment proceeds through a number of avenues, including mass media appeals, schools, word of mouth and the Australian Multiple Birth Association (Hopper, 2002), and enrolment into the registry is possible for all zygosities and ages, irrespective of medical history. Since its establishment, the ATR has been funded by a variety of grants, predominantly by the Australian National Health and Medical Research Council (Hopper et al., 2006). Further details are available in publications by Hopper (2002) and Hopper et al. (2006, 2013).

The current sample consists of 3348 individual twins from the ATR, born between 1972 and 1979, and 476 of their non-twin siblings. At the time of assessment, the average age of the sample was 32.1 years. An overview of the recruitment process is shown in Figure 1.

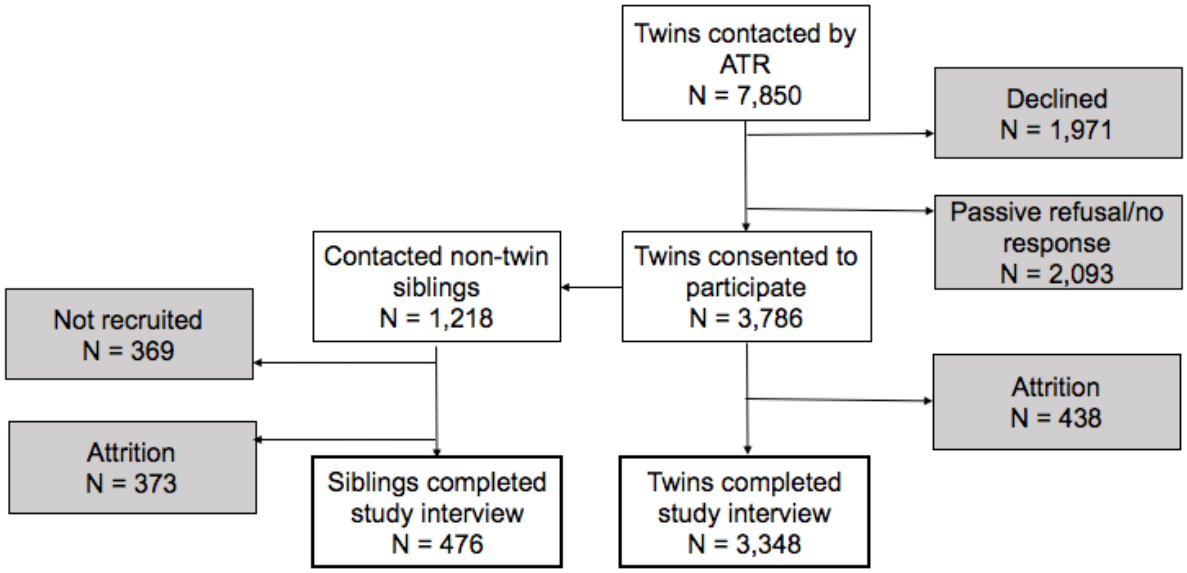

Figure 1. An overview of the recruitment process, based on the description in Lynskey et al. (2012).

Recruitment for the study began in 2005 and was conducted by the Queensland Institute of Medical Research (QIMR). To comply with ATR ethical standards, mail and then telephone contact was established first by the ATR. If individuals were interested in participating after receiving information on study goals and procedures, and were willing to share their contact details, they were contacted by QIMR (Lynskey et al., 2012). QIMR researchers further explained the purpose of the study, and proceeded with enrolment and obtaining informed consent. Subjects were compensated for completing the assessment with a AUD$50 gift card.

Due to smaller than expected enrolment rates of twins, their non-twin siblings were also contacted for recruitment. In total, 7850 individual twins and 1218 non-twin siblings were contacted, of which 3824 individuals, including twins and non-twin siblings completed a full or partial interview. Table 8 summarises the composition of the sample by sex and zygosity group.

Table 8. Composition of sample by sex and zygosity group.

	Female	Male	Total
MZ	976	490	1466
DZ (same sex)	741	373	1114
DZ (opposite sex)	308	438	746
Unknown zygosity	13	9	22
Non-twin siblings	267	209	476

2.2 Ethical approval and funding

The data collection was funded by National Institute on Drug Abuse (NIDA) grant: DA18267 and facilitated through review by the Australian Twin Registry. Twins Research Australia receives support from the National Health and Medical Research Council through a Centre of Research Excellence Grant, which is administered by the University of Melbourne. Ethical approval for data collection was obtained from the Australian Twin Registry, QIMR and Washington University in St Louis. King's College London Research Ethics Subcommittee approved access and storage of the data.

2.3 Measures

Computer-assisted telephone interviews based on the Australian version of the Semi-Structured Assessment of the Genetics of Alcoholism (SSAGA-OZ; Bucholz et al. 1994)) were used to assess twins on a number of mental health-related variables. Interviews were administered by lay interviewers, who underwent two weeks of interview training. With the participant's permission, interviews were recorded for data quality purposes. Members from the same family were assessed by different interviewers, to avoid bias.

2.3.1 The SSAGA-OZ

The SSAGA-OZ is a semi-structured questionnaire, which has been adapted from the original SSAGA (Bucholz et al., 1994) to suit an Australian population and to be

administered through the telephone. The SSAGA-OZ has been widely used in family studies of alcohol dependence and collects detailed information on patterns of DSM-IV (American Psychiatric Association, 2000) symptomatology across a range of mental health and substance use disorders. In the original SSAGA, assessments of these disorders, including DSM-III-R MDD and Cannabis Abuse/Dependence, have been shown to have good reliability (Bucholz et al., 1994) and validity (Hesselbrock et al., 1999). Inter-rater reliability was tested within and across centres, and had a kappa of .82 for Cannabis Abuse/Dependence in both studies (Bucholz et al., 1994). Inter-rater reliability for lifetime MDD had a kappa of .65 in the within-centre study and .74 in the cross-centre study. Validity was tested by comparing the SSAGA instrument to the Schedule of Clinical Assessment in Neuropsychiatry (SCAN), a cross-culturally valid instrument (Hesselbrock et al., 1999). The measures showed good concordance for MDD, with a kappa of .70. Only cannabis dependence, not abuse, was examined here, and, for all positive symptoms assessed on SCAN, had a kappa of 0.71.

The adaptation of the SSAGA interview used for this cohort was enriched for variables related to cannabis involvement. Variables assessed in the interview included drug abuse and dependence (nicotine, alcohol and other illicit drugs), drug-related behaviours, mental health problems, socio-demographic and family environment factors. Those used for analyses in this thesis are further outlined below.

2.3.1.1 Major Depressive Disorder (MDD)

Participants were asked whether they had ever felt a) ‘depressed or down’, b) ‘sad, blue, low, or discouraged’ and c) ‘a lot less interested in most things or unable to enjoy the things [they] usually enjoy’ most of the day and nearly every day, for two weeks or more over their lifetime. They were also asked about whether they ‘were a lot more irritable than usual, or […] found that people or things seem to get on [their] nerves a lot more than usual’ most of the day and nearly every day, for two weeks or more before the age of 18. In total, 1823 participants replied ‘yes’ to any of these four gateway questions, and they were asked further DSM-IV MDD criteria.

For analyses in this thesis, MDD was coded in accordance with the latest DSM, and as a binary variable. Although the criteria used in the interview were from the DSM-IV, there were no significant changes in coding for MDD between DSM-IV and DSM-5. An individual was coded as affected if they met at least five out of nine DSM-5 criteria, including 'depressed mood' or 'loss of interest or pleasure'. Individuals were only coded as affected if they reported that their MDD symptoms did not occur within two months of bereavement and caused a significant impairment in their occupational, social and family responsibilities.

In total, MDD criteria were assessed in five rounds of questions. Out of 1081 individuals who were asked about DSM-IV criteria for MDD, 930 reported criteria in the first round of questions. Individuals were asked about a following episode of MDD if they met criteria in the previous round or if individuals recalled an episode which occurred in a specific situations which met the DSM-IV exclusion criteria (e.g. bereavement). Over five rounds of questions, 968 individuals met the DSM-5 criteria for MDD (see Table 9). Data from 35 participants were missing on all questions on depression.

Table 9. Prevalence of MDD among males, females and overall.

MDD	Males	Females	Total	Percentage
Absent	1122	1699	2821	74.45%
Present	247	721	968	25.55%

The higher prevalence of MDD in females (29.79%) than in males (18.04%) is compatible with findings from general population surveys reporting gender differences in MDD (Kessler et al., 2005, 1994).

2.3.1.2 Cannabis Use Disorder (CUD)

Interviewees were first asked whether they ever had the opportunity to use marijuana or hashish. Out of all participants, data were missing for 27 and 3399 (89.54%) reported that they had an opportunity to use cannabis. In addition, 2601 (68.52%) reported having used cannabis (i.e. marijuana or hashish) at least once in their lifetime. Lifetime frequency of cannabis use was also asked about and participants who reported having used cannabis at least 11 times in their lifetime

were assessed on DSM-IV criteria of lifetime cannabis dependence and abuse. Abuse items included 'hazardous use', 'social/interpersonal problems related to use', 'neglecting major roles to use' and 'legal problems'. Dependence items included 'tolerance', 'using larger amounts or using longer', 'repeated attempts to quit or control use', 'much time spent using', 'physical or physiological problems related to use', and 'activities given up to use'. In addition, an assessment of cannabis withdrawal was available (see Verweij et al. 2013). However, no assessment of 'craving', a criterion introduced in DSM-5, was available for this sample.

CUD was coded as a binary phenotype. To approximate DSM-5 criteria, individuals were coded as 1 ('affected') if they reported at least two symptoms of DSM-5 CUD, except for 'craving'. The remaining participants were coded as 0 ('unaffected'), whether or not they reported a lifetime history of cannabis use. The 'legal problems' criterion was removed from DSM-5 and therefore was not included in the definition of CUD. The prevalence of CUD in the sample was 14.75% (see Table 10).

Table 10. Prevalence of CUD among males, females and overall.

CUD	Males	Females	Total	Percentage
Absent	1073	2164	3237	85.25%
Present	299	261	560	14.75%

The higher prevalence of CUD in males (21.79 %) than in females (10.76%) is also consistent with results from general population surveys (Kessler et al., 2005, 1994).

These estimates are derived from treating all individuals who report CUD symptoms as "affected" and all individuals who do not report symptoms as "unaffected", regardless of their opportunity to use. An alternative approach would have been to only include individuals who had the opportunity to use cannabis in the "unaffected" category and treat all other cases (10.46% of the sample) as missing for CUD. There is no agreed way of deciding between these approaches, either may lead to bias in twin estimates. The best way to reduce this bias is to decide on an approach based on the genetic structure of the phenotypes involved: if a genetic correlation can be expected between opportunity to use and CUD, removing cases which have not had the opportunity to use would unduly disregard genetic information (Heath, Martin, Lynskey, Todorov, & Madden, 2002). In this twin cohort, age of opportunity to use

cannabis and cannabis abuse/dependence have been estimated to overlap genetically to a substantial degree: most of the genetic variance in cannabis abuse/dependence was explained by genetic factors shared with age of opportunity to use (Hines et al., 2018). To reduce bias in genetic and environmental estimates and maximise statistical power, individuals who have not had the opportunity to use were therefore still included in the sample and treated as "unaffected".

2.3.1.3 Covariates

Since the majority of covariates were only used in epidemiological analyses (Chapter 4), details on how these were measured and coded as well as descriptive statistics will be provided in Chapter 4 (Table 13).

2.4 Statistical analyses

This section will focus on the epidemiological analyses (presented in Chapter 4) and the discordant twin analyses (presented in Chapter 5), which were both conducted in STATA (StataCorp, 2015). Twin model analyses which were not based on logistic regression methods will be outlined in Chapter 3.

2.4.1 Epidemiological analyses

The first aim of the epidemiological analyses is to provide unadjusted and adjusted estimates of the association between CUD and MDD, in order to test whether there is a significant co-morbidity in the current sample. Based on previous literature, it was assumed that this co-morbidity would be significant but would become attenuated after adjustment for covariates. If a diminished association was found, the second aim was to examine which covariates may have played a role in this. Previous research has identified a number of variables as likely covariates (see Table 2), which will be matched with variables in the current dataset. These were used as covariates in the epidemiological analyses.

As a test of an unadjusted and adjusted association between MDD and CUD, as well as the effects of covariates, multivariable logistic regression analyses were

considered most suitable. These analyses allowed the accommodation of two important features of the current data: firstly, that MDD and CUD are binary outcomes, as is common in research on psychiatric phenotypes; secondly, that observations in this dataset are not independent. Independent standard errors are a prerequisite for a logistic regression. Given that the sample consisted of twins, and therefore clustered data, the standard errors were not considered independent. Huber-White corrections were used to account for this clustering and to allow logistic regression to be used. The statistical software used for the current analysis, STATA, contains a standardised option for Huber-White corrections (Williams, 2000).

Additionally, most previous literature examining the association of cannabis involvement and depression has used binary phenotypes and reported ORs. ORs were also used to express the increase in outcome likelihood due to each predictor. Conducting multivariable logistic regressions allowed for a direct comparison between current and previous findings.

A remaining challenge for the epidemiological analyses was to reduce the impact of multicollinearity, where two or more predictor variables are highly correlated. Multicollinearity can lead to larger standard errors around the regression coefficients and confidence intervals around ORs, consequently leading to an incorrect assessment of the significance of covariates. Multicollinearity will be checked using the Variance Inflation Factor (VIF) criterion. A higher VIF indicates that a variable shows more evidence of multicollinearity with other variables.

The VIF and polychoric correlations (correlations for ordinal variables) were used as indicators to decide if variables should be combined or excluded. If correlated variables with a high VIF belonged in a similar category (e.g. different variables related to parental problems), a composite variable was created. If the correlated variables were from different categories (e.g. parental conflict and nicotine dependence), the more strongly associated predictor was kept in the model instead of the less highly associated predictor.

2.4.1.1 Multivariable Logistic Regression in STATA

Both CUD and MDD were considered as outcome and predictor variables, since Chapter 4 will discuss the existence of a relationship in either direction when controlling for other variables, rather than presupposing a direction of effect. Multivariable logistic regressions were used to determine the probability of the outcome variable, for example, being affected by MDD, given the values of the predictors, for example age or sex (Cohen, Cohen, West, & Aiken, 2003). Mathematically, the formula for such a regression can be expressed as:

Equation 1. Multivariable logistic regression formula.

$$P(Y) = \frac{1}{1 + e^{-(b_0 + b_1 X_{1i} + b_2 X_{2i} + \ldots b_n X_{ni})}}$$

where P(Y) is the probability of the outcome variable as a function of the logistic transformation of a linear regression equation, which ensures the outcome is bounded by 0 and 1 and therefore a valid probability. This equation estimates coefficients (b_n) for each known value of a variable (X_{ni}), which are a measure of the strength and direction of influence for each contextual variable present in the model.

The coefficients of each predictor variable (b_n) are found by maximum-likelihood estimation, i.e. optimising the joint set of parameters *b*, given the data, with respect to the specified likelihood function of the model. In the case of logistic regression, the likelihood function aims for coefficients which produce the smallest difference between the outcome predicted by the fitted model from the data and the outcome that was actually observed in the data.

The fit of the model is assessed as the difference between these predicted and observed outcomes, which can be calculated through various statistics. The Wald test will be reported, instead of likelihood ratios, since it is recommended when working with clustered data (e.g. twin data; Sribney 2005). A statistically significant Wald statistic means that at least some of the coefficients in the model are significantly different from 0.

The selection of covariates was based on a review of longitudinal studies (see Table 2) reporting on the co-morbidity between depression and cannabis involvement. Covariates which were significantly related to depression, cannabis involvement or both, were summarized into categories (see Chapter 4) to facilitate the matching process with the current dataset. A dataset variable was selected for epidemiological analyses if it matched one of these categories and contained age of onset information.

To provide an 'unadjusted' estimate of the association between CUD and MDD, a multivariable regression between CUD and MDD was computed, which only included sex and age. To obtain an 'adjusted' estimate of the association, and assess the importance of covariates, a multivariable regression was performed first, which provided initial estimates of the ORs for each covariate, showing the results found with the full model that includes all selected variables based on previous literature. Thereafter, backward stepwise regressions were performed to sequentially remove variables which did not pass a predetermined threshold of statistical significance (in this case 0.10, though other values were explored) and thereby identify a more parsimonious model with minimal loss of explanatory power (e.g. *pseudo-*R^2). The models were restricted to include CUD or MDD as predictor variables.

These models were exploratory and not intended to test a particular theory, but rather to reveal which predictors may play an important role in the co-morbidity between CUD and MDD. Backward, rather than forward regressions were selected to avoid suppressor effects. A reduced model also limits the possibility for multicollinearity and is therefore less likely to produce inflated standard errors. Following standard practice, the *p*-values of the covariates were adjusted for the total number of covariates included in the model (the Bonferroni correction). In other words, a *p*-value of .050 reported in a STATA regression model (see Table 17) with 10 predictors would have been reported as .005 (.050/10) without correction.

2.4.2 Discordant twin analyses

Logistic regression analyses were also employed for discordant twin analyses, in which the OR between MDD and CUD was measured in MZ twins discordant for

both disorders. This involves within-family comparisons, so a conditional logistic regression was used in STATA (StataCorp, 2015). Twins from the same family were matched and the likelihood of the outcome variable (MDD), given the predictor variable (CUD) within families, was analysed. Only twins discordant for both phenotypes could be included in the analysis, since within-pair variation is required. Because within-pair comparisons in MZ twins control for all genetic and shared environmental (but not unique environmental) influences, and the sample size was predicted to be small, no further control variables were included in the model. With only MDD and CUD in the model, the OR would have been identical for either MDD or CUD being used as outcomes, so only one outcome (MDD) and predictor (CUD) were used.

3 Twin methodology

The second and fourth aim of the thesis involve applying twin methodology to elucidate the mechanisms underlying the co-morbidity between Cannabis Use Disorder (CUD) and Major Depressive Disorder (MDD). This objective is founded in research, indicating that these phenotypes are co-morbid and both influenced by heritable factors (see Chapter 1). Therefore, genetically sensitive statistical methods are well placed to enrich our understanding of the relationship between these disorders. The following chapter will outline the foundations of the twin methodology, discuss its advantages and limitations, and explain how it can be used to break down the relationship between CUD and MDD into genetic versus environmental factors.

3.1 Introduction

Historically, it has been of great interest to know to what extent human traits are influenced by genetic versus environmental factors. For centuries, various prominent figures, including philosophers, scientists and politicians, have bitterly debated whether differences between humans are best explained by 'nature' or 'nurture' (e.g. Galton, 1874). Recent research has shown that both are equally important (e.g. Polderman et al., 2015) with some traits, disorders and behaviours being influenced more by genetic factors (e.g. Schizophrenia; Sullivan, Kendler, & Neale, 2003) and others more by environmental factors (e.g. Major Depressive Disorder; Kendler, Gatz, Gardner, & Pedersen, 2006). Nowadays, quantitative geneticists often aim to explore complex patterns of genetic and environmental effects within and between specific traits. This chapter will describe how genetically sensitive data can also be used to explore hypotheses of causality and other forms of co-occurrence between disorders.

As mentioned in Chapter 1, twin studies are one of the most popular methods to decompose individual differences – or the variance – within a trait into genetic and environmental contributions. MZ twins share 100% of their DNA, while DZ twins share around 50% of their segregating genes. Therefore, differences in concordance for a trait between MZ and DZ are most likely due to genetics. Using structural equation modelling (described later in this chapter), it is then possible to capitalise on

the differences in genetic resemblance between the twin types, to estimate to what degree traits, and the relationships between them, are influenced by genetic and environmental factors. While not the only genetically informative study design, twin studies have important advantages over family and adoption studies. Compared to family studies, it is possible to distinguish between influences stemming from genetic factors and those from the family environment (Rijsdijk & Sham, 2002). Adoption studies are difficult to conduct, as adopted siblings and other family members can often not be traced. Additionally, the factors which lead to adoption may introduce biases into a dataset (Rijsdijk & Sham, 2002).

The following chapter will outline the ideas and assumptions behind twin studies, and how twin data is used in structural equation models to answer hypotheses about the relative importance of environmental and genetic influences. It will start by explaining how the variance within a population is assumed to be caused by such influences. Then, it will discuss which specific genetic and environmental effects are present and how those may interact. It will outline how the covariance between twin pairs can be utilised to estimate such effects in theory and in practice. The application of the twin method to binary data will be discussed. Moreover, the assumptions and limitations of the twin design will be detailed. Lastly, the chapter will outline types of twin models applied within this thesis: bivariate threshold liability models and Neale and Kendler co-morbidity models.

3.2 Causes of individual differences

Twin studies rest on the assumption that the phenotypic variance within a trait can be decomposed into variance due to genetic and environmental factors. Combined, 'nature' and 'nurture' are assumed to give rise to the normal distribution of individual scores on a measured trait. In support of this assumption, the following section will introduce the polygenic theory of inheritance, which suggests that a multitude of genes influencing human behaviour are inherited in a manner which gives rise to a normal distribution of phenotypic categories. Environmental influences are then thought to blur the differences between these categories and lead to a continuous distribution.

3.2.1 Polygenic theory of inheritance

Many human traits, such as depressed mood, can be described as 'quantitative'. This means that an individual's score on a trait could be placed along a continuum (e.g. a number of days feeling depressed in the past two weeks), which follows a normal distribution. On such a continuum, differences between people are of degree rather than kind. Clinical disorders, such as MDD, tend to be diagnosed in a binary fashion: it is practical for psychiatrists and insurers to keep a binary distinction of 'affected' and 'unaffected'. However, these categories are created on either side of an artificial threshold on a normal distribution. MDD symptoms, such as fatigue or depressed mood, can be measured continuously and studies applying taxometric procedures suggest that individuals tend to differ in their levels of depression in a continuous way, rather than by categories (e.g. Hankin, Fraley, Lahey, & Waldman, 2005; Slade & Andrews, 2005). Previous research also supports that CUD can be conceptualised as a continuous disorder (Baillie & Teesson, 2010).

According to the polygenic theory of quantitative traits, this continuous distribution along which individuals differ on a trait is brought about by the effects of multiple genes in addition to environmental influences (Fisher, 1918; Plomin, Haworth, & Davis, 2009). Genes create variability in a population, because each individual can inherit different combinations of alleles at any genetic locus: each gene can have multiple alleles within a population (e.g. Handsaker et al., 2015) and there are

several possible allele combinations. Since differences in allele combinations correspond to differences in human traits (e.g. Bertolino et al., 2004; Egan et al., 2001), genetic variation at each locus contributes to a normal distribution within a trait (see Figure 2).

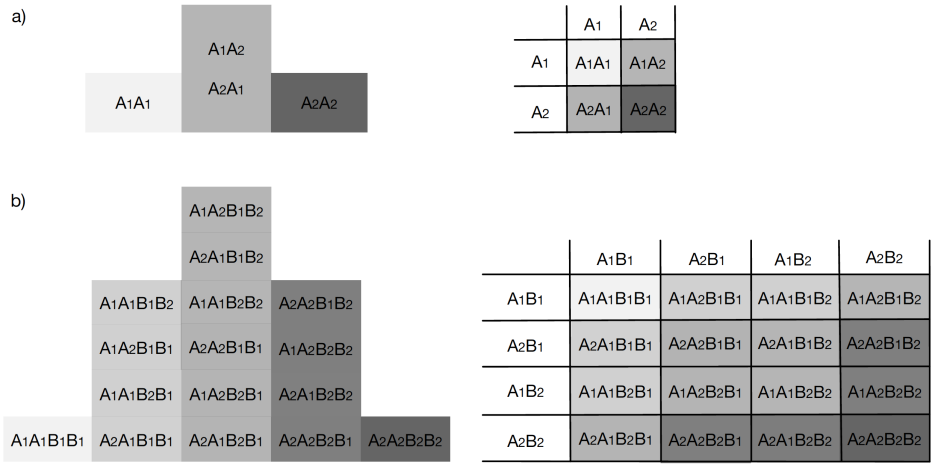

Figure 2. Normal distribution of the phenotype influenced by alleles at two loci. Different phenotypic categories are marked by shading. Equal allele frequencies are implied.

For simplicity, one could imagine that the type 2 allele (e.g. A_2) of each gene contributes to an increase in a trait (e.g. depressed mood), while the type 1 allele (e.g. A_1) contributes to a decrease in this trait. Each combination of alleles gives rise to a qualitatively different phenotype. These phenotypic categories are increased as genetic loci influencing the trait are added (see Figure 2b).

In the current example, individuals with an equal amount of type 1 and 2 alleles are situated around the mean. When allele frequencies are equal, genetic recombination throughout generations gives rise to more heterozygotes (see Figure 2b), i.e. individuals, who inherited two *different* alleles at one genetic locus. Therefore, most people will display medium levels of depressive mood. There will be a minority of individuals inheriting almost only type 1 or type 2 alleles, and therefore displaying very low or very high levels of depression.

Each additional gene that contributes to the phenotype exponentially increases the amount of possible genotypes in the population. This leads to a quasi-continuous normal distribution of phenotypes for quantitative traits, since these traits are estimates to be influenced by thousands of genes (Plomin et al., 2013).

In this example, it is implied that this distribution of phenotypic differences occurs as alleles of different genes combine, segregate and recombine over generations. They do so according to Mendel's laws of inheritance (Plomin et al., 2013): each allele within an individual has a 50% chance of being passed down to the next generation and alleles for different traits assort independently of each other. Although there are some violations of these principles, they do not alter the general principles described above.

In a seminal paper, Fisher (1918) formally demonstrated how Mendelian laws of inheritance give rise to quantitative traits. He also provides further details on the mathematical explanation of this polygenic biometrical model. Fisher (1918) explains that the effect of the environment on the variance in a population is overlaid on the genetic effects and smooths the genetic categories to create a continuous distribution.

3.2.2 Summary

Individual differences in a quantitative phenotype arise due to genetic and environmental effects. Multiple genetic loci and different combinations of alleles create genetic variance. As these alleles are inherited throughout generations, they give rise to phenotypic categories, which resemble a normal distribution. The addition of environmental effects leads to a smoothing of these categories so that the distribution appears continuous.

3.3 Genetic and environmental variance components: A, D, C and E

The previous section outlined how continuous phenotypic variance in a trait is explained by genetic and environmental effects. However, twin studies further sub-divide these effects into more specific sources of variation. These are outlined below.

3.3.1 Environmental and genetic components of phenotypic variance

Broadly, phenotypic variance (σ^2_P) is made up of genetic variance (σ^2_G) and environmental variance (σ^2_E):

$$\sigma^2_P = \sigma^2_G + \sigma^2_E$$

Some component of phenotypic variance is also influenced by the interaction between the two. To be mathematically correct, the equation should therefore be $\sigma^2_P = \sigma^2_G + \sigma^2_E + 2Cov(_{G,E})$. The interaction term at the end of the equation could capture gene-environment correlation (rGE) and/or gene-environment interaction (GxE; Falconer and Mackay, 1996), but these are often omitted for simplicity and parsimony (Purcell, 2013). The implications of this omission will be further discussed below in Section 3.7.

Both genetic and environmental factors can be partitioned further. In terms of environmental variance, we may want to know to what extent a phenotype is influenced by the environment that is shared between relatives (e.g. the home environment) versus their unique environmental experiences (e.g. trauma which only affected one sibling). The shared environment contributes to similarities between twins, while the unique environment contributes to differences. Additionally, genetic effects can be divided into additive and non-additive effects. While examining variance components does not answer which specific environmental and genetic factors contribute to the variance in a trait, it may help to understand which types of variables are likely to most strongly influence the phenotype (Neale & Maes, 2004).

3.3.1.1 Genetic Effects

Variance due to genetic effects can be sub-divided into variance due to additive and non-additive genetic effects. Additive genetic effects are the sum of genotypic effects exerted by any allele at each genetic locus which influence the trait in question (Rijsdijk & Sham, 2002). A genotypic effect is the influence of a genotype on the phenotype. Additive influences account for a large portion of the measurable genetic attributes of a population and the resemblance between relatives (Falconer and MacKay, 1996; Plomin et al., 2013). Statistically, additive genetic effects are expressed by the linear relationship between the number of copies of an allele in a genotype (i.e. gene content) and the genotypic effect (see Figure 3a). If this effect is of equal magnitude for all alleles of a gene, then their influence on the phenotype is said to be 'additive'. Non-additive genetic effects are 'dominance' and 'epistasis'. Dominant genetic effects are present when certain alleles at a locus influence the phenotype more strongly than others. In that respect, dominance reflects the interaction of alleles at one genetic locus (Evans, Gillespie, & Martin, 2002) and would be demonstrated if an increase in the gene content of the dominant allele resulted in a non-linear relationship with the genotypic effect (see Figure 3b). While dominance is the interaction between alleles at the same locus, epistasis is the interaction between genes at different loci (Evans et al., 2002).

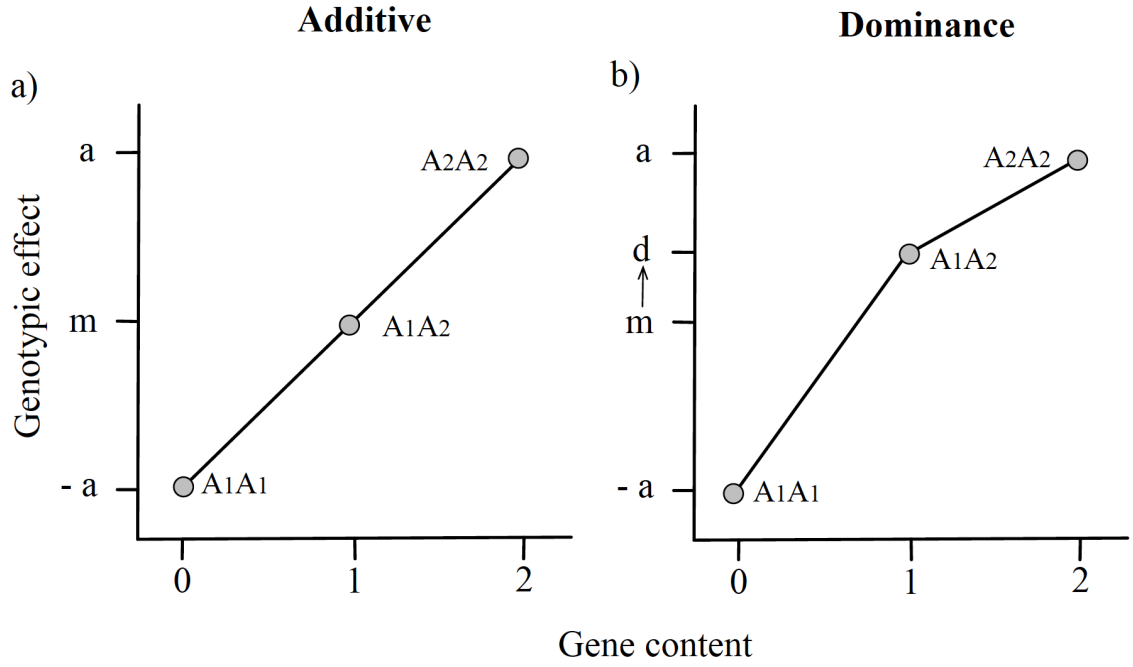

Figure 3a and b. Linear (additive) relationship and non-linear (dominance) relationship between gene content and genotypic effect.

Note. a = average genotypic effect of A_2A_2 (a) and A_1A_1 (-a)
m = midpoint at which genotypic effect is 0
d = dominance deviation

In twin models, additive genetic effects are referred to as A, while dominance effects are referred to as D. As epistasis cannot be measured in twin models, it is not included in the decomposition of genetic variance (see Section 3.7). The total genetic variance contributing to a phenotype is therefore composed of the variance due to additive genetic (σ^2_A) and dominance (σ^2_D) effects:

$$\sigma^2_G = \sigma^2_A + \sigma^2_D$$

If the allele frequencies and genotypic effects are known for a particular locus, it is possible to calculate these variance components. The formula is as follows (Evans et al., 2002):

$$\sigma^2_A = 2pq[(a + d(q - p)]^2$$

$\sigma^2_D = (2pqd)^2$

In this formula *p* and *q* stand for the frequency of both alleles at a locus; *a* stands for the additive genotypic effect and *d* for the effect of dominance. Details on the derivation of this formula are provided by Mather and Jinks (1982), but will not be demonstrated here, as twin models do not investigate the effects of single genes. However, these formula demonstrate several important points (Evans et al., 2002). Firstly, the variance due to additive genetic effects takes into account some contribution of dominance, while the variance due to dominance does not take into account additive genetic effects. Secondly, dominance and additive genetic variance are both influenced by the frequency of alleles. It may therefore occur, for example, that dominance effects are present for a disorder, but are masked by a low frequency of the alleles that exert them (Evans et al., 2002). Lastly, neither formula takes into account the level of genetic variance that arises due to the interaction between different genetic loci (i.e. epistasis). As mentioned, epistasis cannot be assessed in twin studies, but it must be acknowledged as a limitation that a certain degree of genetic variation, which is not captured by dominance or additive genetic effects, is explained by variance due to epistasis (Neale & Maes, 2004).

In twin studies, the proportion of variance in the phenotype attributable to additive genetic effects (i.e. σ^2_A / σ^2_P) is termed 'narrow sense heritability' (Plomin et al., 2013) and abbreviated as h^2, a^2 or A (henceforth h^2). The proportion explained by all genetic effects (i.e. σ^2_G / σ^2_P) is termed 'broad-sense heritability' (Plomin et al., 2013).

3.3.1.2 Environmental effects

Environmental factors are divided into those that make relatives more similar, and those that create differences between relatives. They are sometimes also referred to as the shared and unique environments. In twin models these are abbreviated as C and E, respectively. For instance, the only source of differences between MZ twins who were raised together are their unique environments (E) (Neale & Maes, 2004), as they share 100% of their DNA (A/D) and, by definition, 100% of the environmental influences which contribute to their similarities (C). In contrast to E, C is shared

between family members and contributes to within-family similarities and between-family differences (Neale & Maes, 2004). A study design extending beyond twins, and, for instance, including parents, can provide additional information about the exact sources of shared environmental variation. However, the studies presented in this thesis contain data on twins and non-twin siblings only, and therefore different variance components within C cannot be distinguished. These can include the effect of parents, the environment shared by all siblings, or that shared by the twins only (Neale & Maes, 2004). Environmental variance encompasses both pre-natal and post-natal environmental factors (Neale & Maes, 2004).

Environmental effects are part of the reason why results from different twin studies are cohort-specific (Neale & Maes, 2004). Environmental influences specific to a region or time period may be particularly salient and have a large influence on the phenotype in question, within one cohort, but not another. This may also be more important for some phenotypes than for others. For example, the level of MDD in a population may be particularly responsive to temporary environmental fluctuations (Neale & Maes, 2004). In twin studies, such random effects are included in the estimate of unique environmental variance, which also incorporates measurement error.

3.3.1.3 Total variance composition

Genetic variance is composed of additive and non-additive variance. The latter include dominance and epistasis, but only dominance can be measured in twin studies. Environmental variance can be broken down into variance due to shared and unique environmental influences. The former contribute to similarities within families, the latter to differences. In twin studies unique environmental variance includes a contribution from measurement error.

In total, phenotypic variance can therefore be expressed as:

$$\sigma^2_P = \sigma^2_A + \sigma^2_D + \sigma^2_C + \sigma^2_E$$

This decomposition of phenotypic variance is used in twin studies, but involves some assumptions. As above, any decomposition of variance into additive components mathematically involves interaction terms. The equation above should be as follows:

$\sigma^2_P = \sigma^2_A + \sigma^2_D + \sigma^2_C + \sigma^2_E + 2Cov(_{A,D}) + 2Cov(_{A,C}) + 2Cov(_{A,E}) + 2Cov(_{D,C}) + 2Cov(_{D,E}) + 2Cov(_{C,E})$

However, several assumptions allow us to simplify this equation (Purcell, 2013). Firstly, additive and dominance genetic effects are independent of each other by default, therefore $2Cov(_{A,D}) = 0$. Similarly, there is no interaction between environments which contribute to family resemblance (C) and those which contribute to differences (E), so $2Cov(_{C,E}) = 0$. Secondly, and this is a more contested assumption, genetic and environmental influences are assumed not to be correlated, which equates all covariances involving A or D to 0. The implications of these assumptions will be discussed in Section 3.7.

3.4 Covariance between relatives

Monozygotic (MZ) and dizygotic (DZ) twins can be seen as an elegant experimental design: both share all environmental influences contributing to their similarity (C), while sharing predictably different levels of genetic variance in any trait. MZ twins share 100% of their genetic material and consequently 100% of the additive and dominance genetic variation in a trait. However, DZ twins share, on average, 50% of their segregating genes, 50% of the additive genetic variation in a trait and 25% of the variance in dominance effects. Formal proof for these differences between MZ and DZ twins can be found in Mather and Jinks (1982) and in Neale and Maes (2004). The covariance between twins can be summarised in the following formulae:

$Cov^{MZ}(x,y) = \sigma^2_A + \sigma^2_D + \sigma^2_C$

$Cov^{DZ}(x,y) = \frac{1}{2}\sigma^2_A + \frac{1}{4}\sigma^2_D + \sigma^2_C$

The fact that MZ and DZ twins are only thought to differ in terms of the genetic variance components can be utilised by twin models to estimate these components,

as well as σ^2_C and σ^2_E. The rationale behind this can be demonstrated using a simple calculation, which has been termed Falconer's formula. Using the formulae below, it is possible to approximate the relative importance of different variance components. Rather than estimating the unstandardised variance components, they are expressed as proportions of phenotypic variance. As discussed in the previous section, the proportion of phenotypic variance attributable to additive genetic (i.e. σ^2_A/ σ^2_P) effects is termed h^2. Consequently, other proportions of variance will henceforth be called c^2 (i.e. σ^2_C / σ^2_P), d^2 (i.e. σ^2_D / σ^2_P) and e^2 (i.e. σ^2_E / σ^2_P). For instance, h^2, c^2/d^2 and e^2 could be estimated by the following formulae (Sham, 1998), which are derived by constructing simultaneous equations from MZ and DZ covariances.

$h^2 = 2(r\text{MZ} - r\text{DZ})$

$c^2 = 2r\text{DZ} - r\text{MZ}$

$e^2 = 1 - h^2 - c^2$

or

$h^2 = 4r\text{DZ} - r\text{MZ}$

$d^2 = 2(r\text{MZ} - 2r\text{DZ})$

$e^2 = 1 - h^2 - d^2$

It is important to note that c^2 and d^2 cannot be estimated at the same time using structural equation models (see next section). This is because there are two observed statistics (rMZ and rDZ), from which one can estimate two parameters (h^2 and c^2; Sham, 1998). Thereafter, e^2 can be obtained, because it is known that $h^2 + c^2 + e^2$ equal 1.

Additionally, twin models can only return reliable estimates if rMZ is not smaller than rDZ and never larger than 4 * rDZ (Sham, 1998). If only σ^2_C contributes to twin correlations and $\sigma^2_A = \sigma^2_D = 0$, then both twin correlations should be equal, because

both MZ and DZ twins should share all of σ^2_C. If $\sigma^2_C = \sigma^2_A = 0$ and only σ^2_D contributes to the twin similarity, then the MZ twin correlation can be four times higher, because they are all of σ^2_D, while DZ twins share $\frac{1}{4}\sigma^2_D$.

While a crude estimate of the variance components can be obtained with such calculations by hand, more exact estimates and more advanced hypotheses are examined using structural equation models. For example, it is not possible to use Falconer's formula to examine the genetic relationship between multiple variables, or to calculate confidence intervals around estimates. Nevertheless, Falconer's formula and structural equation models have in common that estimates of variance components are obtained by using simultaneous equations with MZ and DZ twin covariances. The next section describes structural equation models in greater detail.

3.5 Structural equation models

The previous sections have introduced a theoretical account of the sources of variance contributing to a phenotype and how these can be estimated from the covariance between relatives. To formulate these theoretical assumptions and associated hypotheses in a testable way, it is necessary to generate mathematical models (Neale and Maes, 2004).

The path diagram below aims to put the previously introduced theory into a mathematical form (see Figure 4). It hypothesises that phenotypic variance (e.g. Phenotype twin 1) can be decomposed into A, C, D and E (henceforth used instead of σ^2_A, σ^2_D, σ^2_C and σ^2_E). It is further assumed that MZ twins share no E, all of A, all of D and all of C, while DZ twins share no E, $\frac{1}{2}$ of A, $\frac{1}{4}$ of D and all of C. Only ACE or ADE combinations can be modelled at any one time. These variance components are unmeasured, or latent factors, and are therefore represented by circles, while the squared boxed in the diagram represent observed variables. Each latent variable (e.g. A) has a certain magnitude of effect (e.g. a) on the observed variable (e.g. Phenotype Twin 1). This causal relationship is indicated by a single-headed arrow, and the effect (e.g. a) is a partial regression coefficient. Correlations are expressed as double-headed arrows along with a correlation coefficient. By following the path tracing rules (see e.g. Rijsdijk & Sham, 2002), one can obtain the estimated relationship between each connected variable.

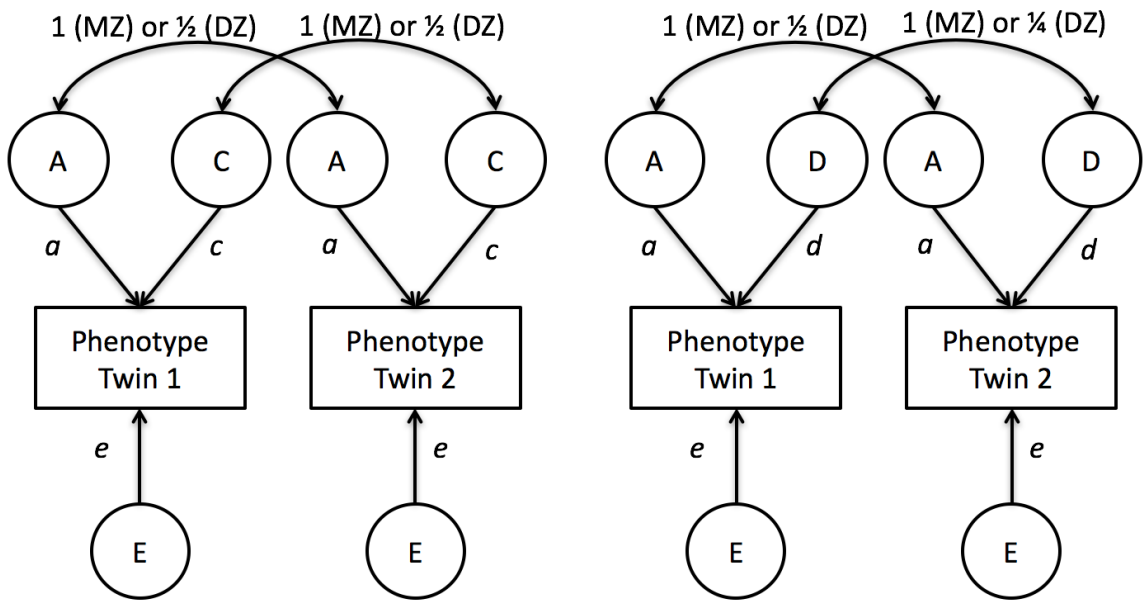

Figure 4. Path diagrams showing variance components and correlations between twin pairs for ACE and ADE models. Partial regression coefficients from latent factors to phenotype are displayed on the single-headed arrows, correlations between latent factors on double-headed arrows.

To analyse twin data and to adequately accommodate all these assumptions several properties are required of a mathematical model: 1) it must estimate latent (i.e. unmeasured) factors, since there is no direct measurement of environmental and genetic effects; 2) it must estimate the causal influence (e.g. via regression methods) of those factors on a trait and the correlation between these factors across traits; 3) It must estimate the conditional relationships of all factors in the model. Structural equation modelling (henceforth SEM) is therefore well suited for this task, as it combines regression analysis, confirmatory factor analysis and path analysis.

In SEM, the phenotypic covariance structure of the data is used to extract latent genetic and environmental factors and estimate the extent of phenotypic variance that is explained by them. MZ and DZ twin covariances can be conveniently expressed in variance-covariance matrices (see Table 11), and be decomposed in terms of the same parameters: A, C, D and E. These parameters need to be estimated in SEM. The aim of the model is to use the minimum number of parameters and simultaneously generate parameter values that predict an expected variance-covariance matrix resembling the observed data as closely as possible (Purcell, 2013).

Table 11. Variance-Covariance matrices for univariate twin analyses.

MZ	Twin 1	Twin 2
Twin 1	Var	
Twin 2	Cov	Var
	Twin 1	Twin 2
Twin 1	A + C/D + E	
Twin 2	A + C/D	A + C/D + E

DZ	Twin 1	Twin 2
Twin 1	Var	
Twin 2	Cov	Var
	Twin 1	Twin 2
Twin 1	A + C/D + E	
Twin 2	$\frac{1}{2}$A + C/$\frac{1}{4}$D	A + C/D + E

During the model fitting process, the statistical program will sample multiple values of A, E and C/D to identify the parameter values which make it most likely to obtain the observed variance-covariance matrix given the model. MZ and DZ data are considered simultaneously, and the parameters must fit the observed variance-covariance matrices of both. This optimisation is part of a method called maximum likelihood estimation (Neale and Maes, 2003),

The highest likelihood obtained by the parameters specified under the model is then compared to the likelihood obtained under a *saturated model*, a model with perfect fit to the data (Rijsdijk & Sham, 2002). This comparison in *goodness-of-fit* is achieved by computing a ratio between the likelihoods, such as the chi-square statistic (Rijsdijk & Sham, 2002). If the difference between the parameterised and unparameterised models is not significant, the ACE or ADE (parametrised) model is assumed to fit well. Further model comparisons can be conducted. For example, it may be worth examining whether a model with fewer parameters (e.g. only A and E) may be more parsimonious. Not all specified parameters may explain the data to a large enough extent. To this end, it is possible to construct nested models, where parameters of one model are fixed to 0 in a sub-model (Purcell, 2013). The best fitting and most parsimonious model can be chosen based on the Akaike Information Criterion (AIC; Akaike, 1987). According to Burnham and Anderson (2002), an AIC difference of three or greater indicates that the model with the lower AIC has substantially more support.

3.5.1 Advantages and limitations of SEM

SEM has several critical advantages for twin data analysis. Maximum likelihood estimation makes it possible to include twin pairs where data is missing, thus avoiding the deletion of cases (Evans et al., 2002) and increasing power. SEM also allows the assessment of complex models and hypotheses and to compare between them (Purcell, 2013). For instance, it is possible to explore the relationship between several phenotypes, examine sex differences and include specific measured variables (Purcell, 2013). Moreover, relationships between factors can be conveniently expressed in the form of path diagrams.

However, a well-fitting model only implies that the hypothesis is plausible, and not necessarily true. There are several models which may have an equally good fit while testing different hypotheses (MacCallum, Wegener, Uchino, & Fabrigar, 1993). Furthermore, an accurate model may not be testable given the available data. For instance, dominance effects are likely to be present but cannot be estimated alongside shared environmental factors (Purcell, 2013). Lastly, there are several underlying assumptions of the model fitting procedure. It is assumed that the components of the model influence each other in a linear fashion, that they additively predict the phenotype and that the examined distributions have multivariate normality (Purcell, 2013). The extent to which these assumptions can be tested is limited.

3.5.2 OpenMx

While structural equation modelling can be performed in various statistical packages, most twin models are conducted using OpenMx (Neale et al. 2016), which is a package for R statistical software (R Core Team 2014). The advantage of OpenMx over other SEM packages includes the fact that it is open source, has been designed to handle large datasets and can be used across different computing environments (Boker et al., 2011).

3.6 Application of twin models to psychiatric phenotypes

From the previous sections it is evident that the covariance between relatives plays a central part in twin models. SEMs take advantage of the covariance structure to estimate environmental and genetic effects. When dealing with continuous data, the concept of covariance is straightforward. Continuous phenotypes contain a variety of values, from which it is possible to generate the exact mean and variance for each sample and sub-sample. From these parameters it is then possible to calculate the covariance between relatives. However, all psychiatric phenotypes analysed in this thesis are defined as binary variables, similarly to previous research reviewed in Chapter 4. Calculating covariance is more complicated in these cases, and will be described in the following sections.

3.6.1 Binary data

Although psychiatric phenotypes (i.e. traits or disorders) are often measured as binary variables (0 = affected, 1 = unaffected), this does not necessarily correspond to the way in which individuals differ in these phenotypes. For example, an individual is considered to be 'affected' by MDD if they have experienced five or more symptoms of depression (see DSM-5; American Psychiatric Association, 2013) for over two weeks, including depressed/irritable mood or loss of pleasure. This disease status can be regarded as binary or further sub-divided into three 'affected' categories depending on the number of displayed symptoms: 'mild', 'moderate' and 'severe'. However, as was mentioned above, individuals tend to differ in their levels of depression in a continuous way, rather than by categories (e.g. Hankin, Fraley, Lahey, & Waldman, 2005; Slade & Andrews, 2005). Furthermore, it can be assumed that psychiatric disorders are polygenic (Wray et al., 2014). In accordance with the polygenic theory of inheritance, it is therefore likely that the risk, or liability, to develop a psychiatric phenotype is continuously distributed as well.

To marry the binary nature of the measure and the assumption that the underlying liability is continuously distributed, the liability-threshold model was proposed (Pearson & Lee, 1901; Sham, 1998). Figure 5 shows a liability-threshold model for binary data. While the underlying liability is thought to be continuous, individuals

above a certain threshold (t) are labelled as ‘affected’. The area under the curve describes the probability of being affected by the disorder, which is influenced by the number of risk factors. This is a mathematical expression of the notion that risk factors for polygenic disorders do not determine their presence, but they influence the likelihood of this (Rutter, Moffitt, & Caspi, 2006)

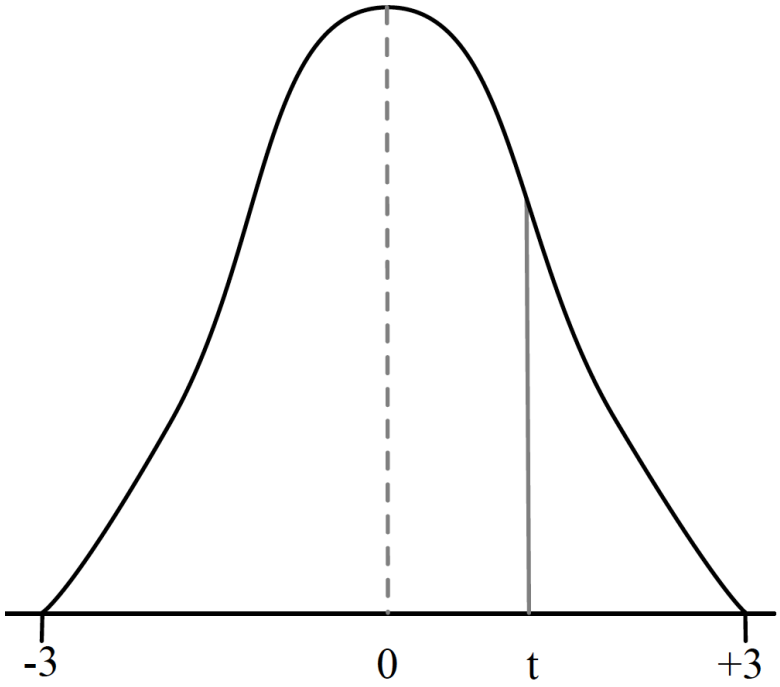

Figure 5. Liability-threshold model for binary data.

As the liability distribution underlying the binary phenotype cannot be observed, this distribution needs to be estimated. There are three parameters which are used to describe such a liability distribution: a mean ($\bar{x}$), a variance (σ^2) and a threshold (t). Because binary data only contain one observed statistic – the probability of being affected by a disorder – only one of the three parameters can be estimated. Identifying restrictions need to be applied to the other two. For twin models, it can be most useful to estimate the threshold, while the variance is fixed to 1 and the mean is fixed to 0 (Purcell, 2013, Sham, 1998). This has the main advantage that the resulting distribution is standard normal, and has well-defined mathematical properties, which simplifies calculations during the twin modelling process.

3.6.1.1 Tetrachoric correlations

The distribution of a given phenotype in one population can be described with a *uni*variate standard normal distribution containing one threshold (see Figure 5). There are only two categories: 'affected' (1) and 'unaffected' (0). Each population has its own distribution and threshold. When examining the correlation between twins (e.g. between MZ T1 and MZ T2), two distributions, and two thresholds, need to be combined. This results in a joint *bi*variate normal distribution with four categories (T1 = 0, T2 = 0; T1 = 0, T2 = 1; T1 = 1, T2 = 0; and T1 = 1, T2 = 1), which represent the different combinations of disorder status (see Figure 6). The areas occupied by these four categories within the bivariate normal distribution represent the proportions of individuals in those categories.

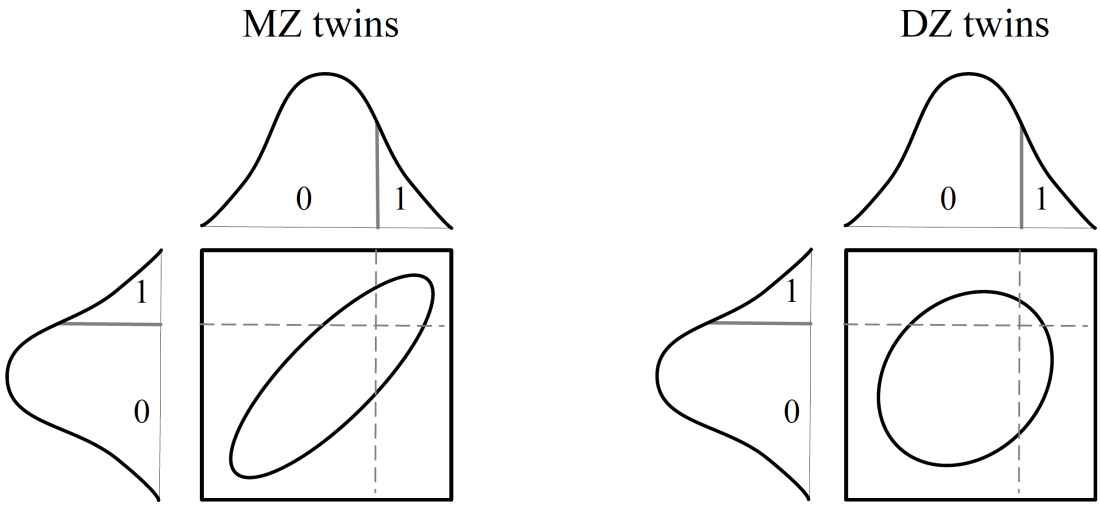

Figure 6. Tetrachoric correlations between twin pair populations. A slimmer ellipse and a stronger correlation are observed on the left, compared to a wide ellipse and a weaker correlation on the right. MZ twin correlations are usually expected to be higher than DZ correlations.

The more twins are within the same disorder status category (T1= 1, T2 = 1 or T1= 0, T2 = 0), the higher the correlation between twins. Visually, the degree of resemblance between relatives can be gleaned from the 'fatness' of the ellipse (Uebersax, 2015): the slimmer the ellipse, the more twin scores correspond to each other (T1 = 0, T2 = 0 and T1 = 1, T2 = 1), and the higher the correlation between

twin 1 and twin 2 samples. The 'fatness' of this ellipse is the so-called tetrachoric correlation (Uebersax, 2015) and is called rho (r*). Tetrachoric correlations were first introduced by Pearson (1900). Because the variances are 1 and therefore standardised for all twin populations, calculating the correlation is equal to calculating the covariance (Purcell, 2013).

Figure 6 shows visual examples of how ellipses and correlations might be expected to differ between MZ and DZ twin pairs. While thresholds are expected to be the same, more MZ twins are expected to fall into corresponding categories, show a slimmer ellipse and therefore a higher correlation.

3.6.1.2 SEM with binary data

In OpenMx, maximum likelihood estimation is used to find values for the twin 1 threshold, twin 2 threshold and r*. The aim is to estimate the parameters in a way that minimises the differences between the expected proportion of twins in the four categories, and in the one that is observed. Tetrachoric correlations are estimated separately for MZ and DZ twin populations. Once correlations have been obtained they are used in simultaneous equations to obtained estimates for h^2, c^2 and e^2.

One assumption made in calculating tetrachoric correlations using the bivariate normal distribution is bivariate normality (Sham, 1998). In other words, it is assumed that the underlying distribution is not significantly different from a normal distribution. While this assumption cannot be tested using binary data, the additional degrees of freedom afforded by adding further thresholds allows for this assumption to be tested (Sham, 1998).

3.7 Assumptions and limitations

The theoretical and mathematical concepts underlying twin modelling rely on a set of assumptions. Some have been mentioned in previous sections and will be outlined here in detail. None of the assumptions and accompanying limitations preclude twin analyses. However, it is important to be aware of them in order to accurately interpret results.

3.7.1 The Equal Environments Assumption

Increased similarity between MZ twins is assigned to genetic factors, because i) MZ twins share all of their genetic material, while DZ twins share approximately half, and ii) MZ and DZ twins are thought to be influenced equally by environmental factors (Rijksdijk & Sham, 2002). The latter statement is frequently referred to as the Equal Environments Assumption and states, in more general terms, that the environmental influence on similarity within a twin pair on a particular variable is independent of zygosity. This has been subject of considerable debate. MZ twins look more identical, are more likely to spend more time together, to have the same friends, to be treated alike by their parents, to visit the same classes and to dress similarly (e.g. Horwitz, Videon, Schmitz, & Davis, 2003; Loehlin & Nichols, 1976; Morris-Yates, Andrews, Howie, & Henderson, 1990).

Critics have argued, therefore, that the environment of MZ twins may contribute more to their intra-twin similarity than the environment of DZ twins (Pam, Kemker, Ross, & Golden, 1996; Richardson & Norgate, 2005). Nevertheless, increased MZ similarity for certain environmental factors does not, in itself, present a violation of the EEA, since these may not impact the phenotype in question (Medland & Hatemi, 2009). Furthermore, it is possible that increased environmental similarity between MZ twins may also be driven by increased genetic similarity (Lytton, 1977), in which case the variance component would be correctly assigned.

With respect to psychological traits and disorders, several studies have attempted to estimate to what degree the Equal Environments Assumption holds true and twin model estimates are accurate. Although evidence is mixed (e.g. Clifford, Hopper, Fulker, Murray, & Rao, 1984; Horwitz et al., 2003), most studies do not find that evidence contraindicatory to the Equal Environments Assumption (Bouchard, Lykken, McGue, Segal, & Tellegen, 1990; Kendler, 1983; Kendler, Neale, Kessler, Heath, & Eaves, 1993, 1994; Loehlin & Nichols, 1976; Matheny, Wilson, & Dolan, 1976; Morris-Yates et al., 1990; Plomin, Willerman, & Loehlin, 1976; Scarr & Carter-Saltzman, 1979; Xian et al., 2000). In terms of MDD and cannabis dependence, no specific violations of the Equal Environments Assumption have been found (Kendler et al., 1993, 1994; Lynskey et al., 2002; Xian et al., 2000).

3.7.2 Assortative mating

When individuals make their partner choices non-randomly, this is termed assortative mating. Non-random, in this context, means that partners would be found to share the same phenotype more often than expected by chance (Falconer & Mackay, 1996). This could be due to environmental or genetic influences (Falconer & Mackay, 1996), but for twin models it becomes particularly problematic if individuals select partners who are genotypically more similar to them. Twins models assume that DZ twins share, on average 50% of A. However, increasing the parental genotypic similarity means that DZ twins might share a higher amount of their segregating genes. This, in turn, would lead to a higher correlation between DZ twins, while leaving the MZ correlation unchanged. If this increase in DZ genetic similarity is left unaccounted for, the heritability of a phenotype would be underestimated (Rijsdijk & Sham, 2002).

There is evidence for increased partner similarity and therefore assortative mating for MDD (e.g. Maes et al., 1998; Mathews & Reus, 2001) as well as CUD (e.g. Merikangas et al., 2009). However, it is unclear to what extent this is driven by genetic or environmental factors, and there is also the possibility that this similarity has developed due to social interaction *after* partner choice (Neale & Maes, 2004). Furthermore, Maes et al. (1998) estimate that the bias in twin studies for psychiatric disorders stemming from partner correlations would be minimal.

3.7.3 Genotype-environment interplay

In twin models it is usually assumed that individuals experience a random sample of environments (Neale & Maes, 2004) and that these are not correlated with genetic effects. Therefore, the formula to partition the phenotype $\sigma^2_P = \sigma^2_G + \sigma^2_E + 2Cov(_{G,E})$ is condensed to $\sigma^2_P = \sigma^2_G + \sigma^2_E$. However, individuals can create environments for themselves and share environments with related individuals, which leads to scenarios where genes and environments are correlated. Moreover, particular genotypes seem to modify an individual's response to the environment, and therefore lead to genotype-environment interaction. These violations of a straightforward partition of σ^2_P into $\sigma^2_G + \sigma^2_E$ are outlined below.

While the examples will refer to one phenotype at a time, the interplay between genotypes and environments can occur both within and between phenotypes (McAdams et al., 2014).

3.7.3.1 Genotype-environment correlation (rGE)

1.1.1.1.2 Passive genotype-environment correlation

Passive rGE occurs because individuals share their home environment with people they are genetically related to (Neale & Maes, 2004). As an example, an individual's depression levels could be influenced by the genes they inherit from their parents or a home environment that is conducive to depression (e.g. Murray & Cooper, 1997). This home environment may, in turn, be influenced by the parental genetic predisposition to depression (Neale & Maes, 2004). Therefore, the genes and environment that children receive from their parents could correlate for this particular phenotype ($2Cov_{(G,E)}$), and seem to do so for mental health problems in general (Rutter & Quinton, 1984). Not taking this interaction into account will lead to biased estimates for the variance components. A positive correlation between the genotype and family environment will overestimate the influence of the family environment (Rijsdijk & Sham, 2002), whether it makes twins more alike or different (Rutter et al., 2006).

If individuals were adopted randomly, passive rGE related to the parental environment does not occur and can therefore be examined in adoption studies. Twin studies, which include non-adopted MZ and DZ twins and no further information on specific genes or environmental influences, do not control for passive rGE.

Passive rGE can also be the result of sibling effects, as twins share genetic material, as well as their mutually influenced environment. Siblings can create environments for each other which can cause resemblance or dissimilarity. Such sibling effects may differ between MZ and DZ twins and increase the difference between both twin types to non-twins (Neale & Maes, 2004). The latter may affect the generalisability of twin data.

1.1.1.1.3 Active genotype-environment correlation

In an active rGE, individuals actively invoke environments based on their genotype (McAdams et al., 2014). In other words, individuals will prefer environments based on their genotype when given the choice of multiple environments (Neale & Maes, 2004). The degree of this choice may vary across cultures and circumstances and be one reason why heritability estimates vary by cohort.

Cross-sectional twin studies are unable to differentiate between active rGE and other genotypic effects, which is why heritability estimates will be increased by positive correlations and decreased by negative correlations (Neale & Maes, 2004). MZ twins would be more likely to choose similar environments based on their increased genetic similarity if there is a positive active rGE, which would further contribute to an increased resemblance between them, and therefore a higher heritability estimate than that of DZ twins.

3.7.3.2 Genotype-environment interaction (GxE)

A genotype-environment *interaction* occurs when an individual's genotype determines their response to the environment. This means that individuals can experience the same environment but react to it differently depending on their genotype. In the case of MDD, several studies have reported that certain genetic variants moderate the influence of stressful life events on the development of MDD (Caspi et al., 2003; Eley et al., 2004; Kendler, Kuhn, Vittum, Prescott, & Riley, 2005). There are also likely to be genotype-environment interactions between cannabis-related genotypes and environmental factors, at the least because environmental exposure to drugs is necessary to develop CUD (Heath, Todorov, et al., 2002). However, molecular GxE analyses require cannabis-related genetic targets, and those are still in the process of being identified (Stringer et al., 2016).

In twin studies, one can assess whether specific environments interact with genetic liability to a disorder by examining twins discordant for that disorder. If the co-twin of an affected twin (e.g. has MDD) is more likely to react (e.g. develop MDD) to an environment (e.g. stressful life event) than the co-twin of an unaffected twin, this

suggests GxE (see Kendler et al., 1995). One study showed that twins with a high genetic risk of developing MDD were more likely to do so following stressful life events than those at low risk (Kendler et al., 1995). However, when specific environments are not measured, GxE cannot be examined.

If an interaction between the genotype and non-shared environment is left unaccounted for, this will be confounded with and lead to an inflated estimate of E (Heath, Todorov, et al., 2002; Rijsdijk & Sham, 2002), since the interaction does not contribute to twin co-variance (Heath, Todorov, et al., 2002). Interaction between the genotype and shared environment, on the other hand, is confounded with and will lead to inflated estimates of A (Heath, Todorov, et al., 2002; Rijsdijk & Sham, 2002), because MZ and DZ twins share the same amount of GxC interactions (100% for MZ and 50% for DZ twins), as genetic effects (Heath, Todorov, et al., 2002).

3.7.4 Gene-gene interaction: epistasis and dominance

In order to exert their effect, genes need to be transformed into proteins. The DNA within a gene is transcribed into a particular mRNA sequence, translated into a chain of amino acids and folded into proteins. These are the building blocks of the human body and can influence behaviour. This sequence of events, between DNA and proteins, can be affected by several factors which determine when, where and how it takes place. Epistasis is the influence of DNA segments in one location on the expression of DNA segments in another location, whether they do or do not code for proteins (Rutter et al., 2006). Dominance is such an interaction at the same genetic locus. Both are genotype-genotype interaction effects and, when present but not taken into account in a twin model, lead to the overestimation of heritability (Mackay, 2014). This is, because MZ twins share more dominance and more epistasis effects than DZ twins, and therefore both are likely to contribute to an increased MZ similarity. However, in a standard ACE model, where neither is taken into account, this increased similarity will be attributed to A. Disorder-specific (e.g. in MDD; Schott et al., 2014) and non-specific epistasis are present and may explain difference between MZ and DZ twins, but these cannot be estimated in a twin model. Dominance can be estimated, and is included in twin models instead of C when the correlation between MZ twins is over twice the correlation between DZ twins. This

has the disadvantage that shared environmental effects are not estimated. Furthermore, it is likely that dominance contributes to an increased similarity between MZ twins even when their correlation is not twice the DZ correlation. Due to these limitations, it is important to remember that A in an ACE model is an estimate of broad-sense heritability, and includes all genetic effects, not just those inherited from the parents.

3.7.5 Importance of assumptions and limitations

The aforementioned assumptions and limitations are critical to take into consideration when interpreting twin study findings. However, the likelihood of violating these assumptions depends on the phenotypes studied. From previous literature, there is no reason to assume that applying twin model analses to MDD and CUD is likely to violate core assumptions of the twin design. As reviewed in Chapter 1, several twin studies have been conducted on clinical definitions of both depression and cannabis involvement.

3.8 Application to MDD and CUD

As previously stated, the overall objective of the thesis is to gain a clearer understanding of the co-morbidity between MDD and CUD by means of twin methodologies. This chapter illustrated that this methodology is suitable to the phenotypes in question, as it is suitable to psychiatric and quantitative phenotypes in general.

The risk of developing MDD or CUD is likely to be continuous and influenced by multiple genes (see Section 3.2.1). Therefore, the variance of MDD and CUD in the population, as well as the covariance between them, can be decomposed into genetic and environmental factors.

Having introduced the theoretical background, the following section will give an overview of the specific twin models applied to answer the research questions explored in this thesis:

1. **Bivariate threshold liability models** will be used to obtain a clear picture of the extent to which environmental and genetic factors influence the variance within and covariance between MDD and CUD.

2. **Co-morbidity models** by Neale and Kendler (1995) will be fit to examine 13 possible aetiological mechanisms through which genetic and environmental factors may give rise to the co-morbidity between MDD and CUD.

3.8.1 Bivariate threshold liability models

Bivariate twin models aim to decompose the relationship between two variables into A, C/D and E, and will be discussed in Chapter 5. Estimates of the partial regression coefficients *a, c/d* and *e* indicate the relative impact of each latent factor on the variance in a single phenotype. Estimates of the correlation between the latent factors, rA, rC/rD and rE, indicate the impact of latent factors on the covariance between phenotypes.

Since models are fit to binary definitions of MDD and CUD, they are bivariate threshold liability models. This means that the decomposition of the variance of each phenotype into genetic and environmental factors is applied to the estimated univariate liability distributions of these phenotypes (illustrated in Figure 5), while the covariance decomposition is applied to the estimated *bivariate* liability distribution (illustrated in Figure 6). The path diagram summarising the relationship between two phenotypes for the whole study population (rather than by twin types, as shown in Figure 4) can be found in Figure 7.

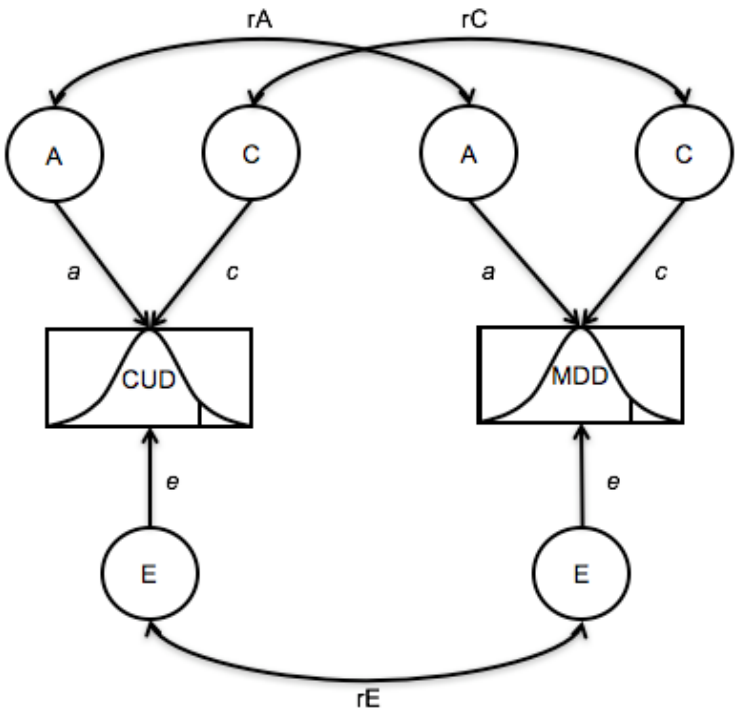

Figure 7. Bivariate threshold liability model for MDD and CUD. Partial regression coefficients are marked on single-headed arrows. Latent factor correlations are marked on double-headed arrows. The ACE model is used as example.

As illustrated in Figure 7, bivariate models fit in this thesis are specified according to the correlated factors solution, which makes no assumption about the order in which individuals experience MDD or CUD. A correlation is estimated between genetic and environmental latent factors for both phenotypes (i.e. rA, rC/rD and rE), which can then be expressed as a proportion of the total phenotypic correlation, indicating the relative importance of those factors in the covariance between disorders.

3.8.1.1 Examining sex differences

Extensions of the bivariate threshold liability model outlined in Neale, Røysamb and Jacobson (2006) can be used to examine whether sex differences are likely to be present and therefore need to be accounted for in the model. Two types of sex differences may present in the data: quantitative and qualitative.

Quantitative sex differences exist when a phenotype is influenced by the *same* genetic and environmental latent factors in males and females, but to a different degree. Allowing separate regression and same-sex correlation paths for males and females can test the importance of quantitative sex differences. A model with separate paths (see Figure 8) is compared to a model where regression and same-sex twin correlation paths are equated across sexes (see Figure 7). If model fit significantly deteriorates when regression paths and same-sex correlations are equated, quantitative sex differences need to be allowed for in the model.

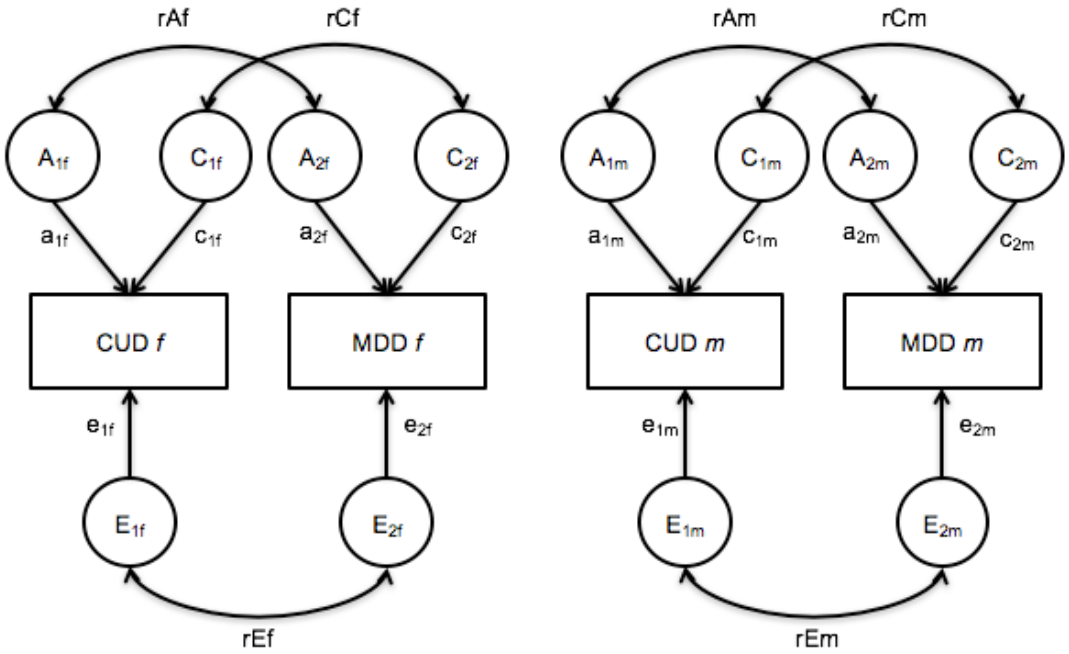

Figure 8. Bivariate threshold liability model with separate paths for males and females. The ACE model is used as example.

Qualitative sex differences are said to be present if males and females are influenced by *different* latent factors. These sex differences can only be estimated

using the covariance between opposite-sex DZ (DZos) twins. If different latent factors influence a phenotype in each sex, then the correlations of those latent factors should be different in DZos twins from those one would usually specify for same-sex DZ twins. In other words, between-twin correlations for A, C and D would be significantly different from 0.5, 1 and 0.25, respectively.

This can be tested by leaving latent factor correlations between DZos twins to be freely estimated and then comparing this specification to a model where they are fixed. If freely estimated DZos correlations significantly differ from those usually assumed for DZ same-sex twins, the fixed model should fit significantly worse and qualitative differences need to be taken into account in the model. Qualitative sex differences are tested separately for A, and C or D, as not enough information is present in the model to estimate more than one at a time. Since E does not contribute to the covariance, one can only test for qualitative sex differences in A,C or D.

Further details on the rationale to examine sex differences and the results of these analyses will be found in Chapter 5.

3.8.2 The Neale and Kendler co-morbidity models

Co-morbidity models based on the work of Neale and Kendler (1995) and Klein and Riso (1993) examine 13 forms of co-morbidity, which were introduced in Section 1.4.3 (Chapter 1), are described in detail in Chapter 6 and summarised in Table 12. Each model tests different assumptions about aetiological processes which may give rise to co-morbidity between two multifactorial disorders (i.e. disorders such as MDD and CUD, which are influenced by multiple genetic and environmental factors). They also provide information on the importance of A, C/D and E. The description in this chapter will focus on the mathematical approach to the co-morbidity models.

Table 12. Thirteen Neale and Kendler (1995) co-morbidity models and the questions these models aim to address.

Model	Question
Single liability model	
1. Alternate Forms	Alternate forms of the same disorder?
Independent liability model	
2. Three Independent Disorders	Co-morbid form is an independent disorder?
Multiformity models	
3. Random Multiformity	Abruptly increase symptoms of each other?
4. Random Multiformity of MDD	MDD abruptly increases CUD symptoms?
5. Random Multiformity of CUD	CUD abruptly increases MDD symptoms?
6. Extreme Multiformity	CUD and MDD abruptly increase symptoms of each other in extreme cases?
7. Extreme Multiformity of MDD	MDD abruptly increases CUD symptoms in extreme cases?
8. Extreme Multiformity of CUD	CUD abruptly increases MDD symptoms in extreme cases?
Correlated liabilities models	
9. Correlated Liabilities	Liabilities are correlated?
10. Reciprocal Causation	CUD and MDD cause each other?
11. MDD causes CUD	MDD causes CUD?
12. CUD causes MDD	CUD causes MDD?
13. Chance	Co-morbid due to chance?

3.8.2.1 Co-morbidity model mathematical approach

To illustrate the logic of the Neale and Kendler (1995) co-morbidity models, two models will be used: The alternate-forms model and the random multiformity model. The alternate-forms model assumes that two disorders (Disorders A and B in Figure 9) are alternative expressions of the same liability distribution. After crossing a common threshold on this shared liability distribution (the probability of crossing the threshold is *U*), a proportion of individuals will experience disorder A (with the probability of *p*), others will experience disorder B (with the probability of *r*) and some will be co-morbid for both (with the probability *p*r*). The probabilities of an individual

not experiencing disorders A and/or B are 1 – *p* and 1 – *r,* respectively, when the threshold has been crossed. The probability of not having both disorders is 1 if the threshold has not been crossed (the probability of not crossing the threshold is *L*).

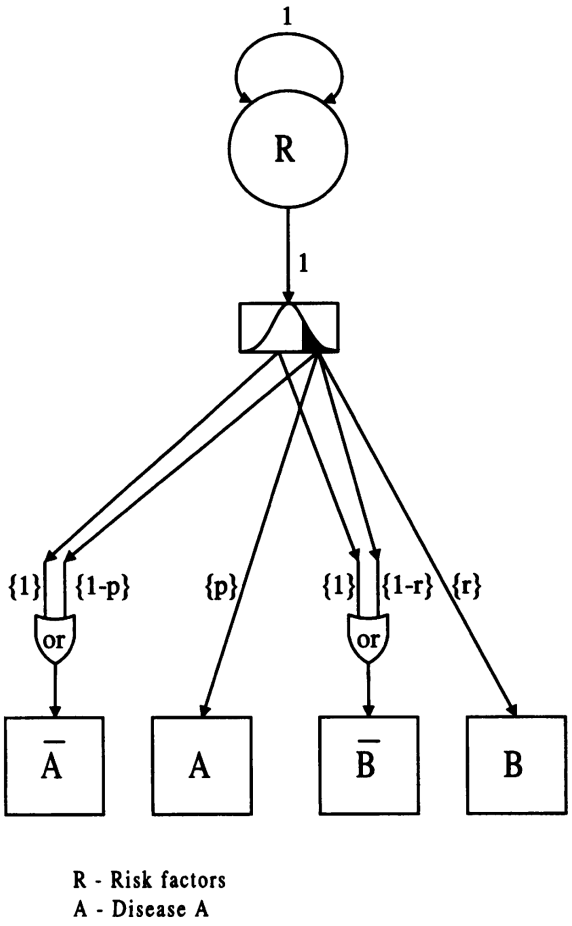

Figure 9. Alternate-forms model schematic from Neale and Kendler (1995). The figure describes the probabilities of a single individual to have various disease statuses. The probability of having disorder A and B are expressed as p and r, respectively. R stands for the latent genetic and environmental risk factors affecting the shared liability of both disorders. Dashes above the letters denote being unaffected by a disorder.

The likelihood of any disease status for an individual under the assumptions of the model can therefore be given in four equations, where P(A,B) is the likelihood of being affected by both disorders:

$P(\underline{A},\underline{B}) = L + (1 - p)(1 - r)U$

$P(\underline{A},B) = r(1 - p)U$

$P(A,\underline{B}) = (1 - r)pU$

$P(A,B) = prU$

In twin models, these equations are expanded to include the probability of disease statuses for two individuals (e.g. P(A1,B1,A2,B2)). Actual rates of individuals within twin disease categories are compared to the ones estimated given the assumptions of the model to determine model fit.

Models can be compared with each other, because the calculations for disease status combinations between twins differ. If, for instance, the random multiformity model were a better fit to the data, the disease status likelihoods for an individual would be better expressed as:

$P(\underline{A},\underline{B}) = L_A * L_B$

$P(\underline{A},B) = (1 - r)\, L_A * U_B$

$P(A,\underline{B}) = U_A * (1 - p)\, L_B$

$P(A,B) = U_A * (U_B + pL_B) + rL_A * U_B$

This would mean that two separate liabilities for each disorder better account for the co-morbidity status of an individual (see Figure 10). All individuals who cross the threshold on the disorder-specific liability develop the disorder and a proportion of those also develop a co-morbidity. Here, *p* and *r* stand for the probability of developing Disorders A and B, respectively, when crossing the threshold on the liability of the other disorder. Under the *Random Multiformity* model, liabilities for both disorders are unrelated, but developing one disorder abruptly increases the chances of developing the other.

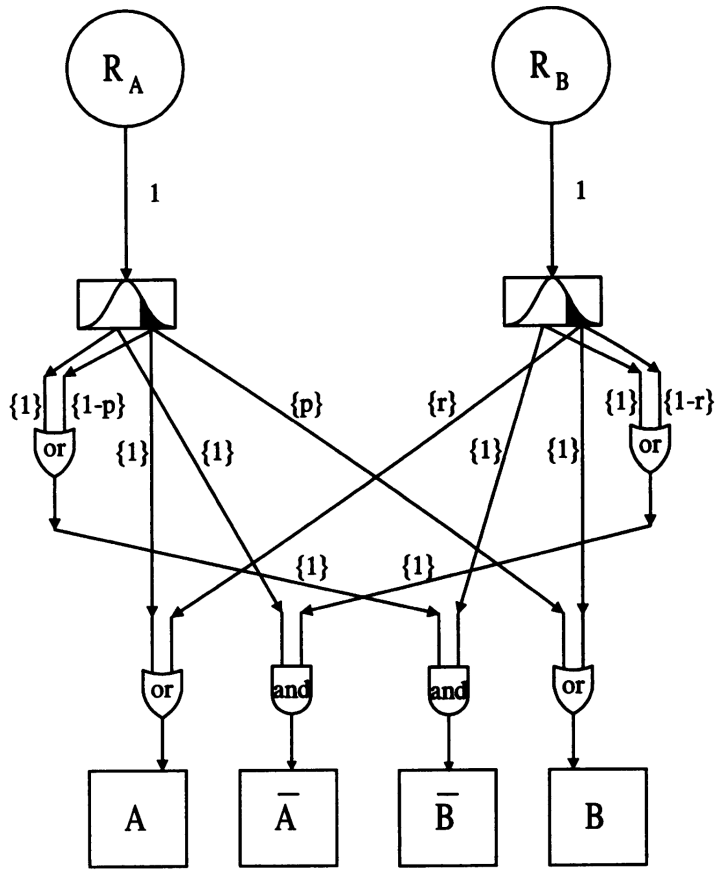

Figure 10. Random Multiformity model schematic from Neale and Kendler (1995). The figure describes a single individual's probabilities of having various disease statuses. The probability of having Disorder A and B are expressed as p and r, respectively. R_A and R_B stand for the latent genetic and environmental risk factors affecting the separate liabilities of both disorders. Dashes above letters denote being unaffected by a disorder.

All models have differing formula for all possible co-morbidity statuses of twins. The assumptions of the best fitting model about the number of liabilities, thresholds and directions of causation allow the researcher to make inferences about the aetiology of the co-morbidity between Disorder A and B when model fits are compared.

3.9 Summary

Individual differences on quantitative traits are influences by genetic (A and D) and environmental (C and E) factors. Twin model analyses – using structural equation models – decompose the variance of a trait into these factors. Twin model assumptions and limitations need to be taken into consideration when interpreting results. MDD and CUD are quantitative phenotypes, i.e. influenced by large number of genes, and twin models can be used to estimate in how far their variances and covariance are influenced by genetic and environmental factors. Since both disorders are measured as binary variables, threshold liability models are the foundation for both bivariate and co-morbidity models, which will be presented in Chapters 5 and 6, respectively.

4 Epidemiological Analyses

4.1 Introduction

The literature reviewed in Chapter 1 presents convincing evidence that cannabis involvement and depression co-occur. Most cross-sectional studies have found a significant association between various measures of cannabis involvement and depression (Degenhardt et al., 2012), while longitudinal studies observed the strongest links (aORs ranging from 1.78 (95% CI = 1.17 – 2.71) to 4.00 (95% CI = 1.23–12.99)) between clinical levels of both (e.g. Bovasso, 2001; Marmorstein & Iacono, 2011; Pacek, Martins, & Crum, 2013).

However, few cross-sectional and longitudinal studies so far have been conducted on MDD and CUD specifically, and many analyses on clinical levels of cannabis involvement and depression have not examined an extensive range of potential covariates. This may be explained by analyses being conducted with existing data which had not been collected to examine cannabis-related or depression-related variables originally. Incidentally, the longitudinal studies finding the strongest ORs between clinical levels of cannabis involvement and depression are among those that control for the fewest covariates (see study summaries in meta-analysis by Lev-Ran et al., 2014 and Pacek et al., 2013). Consequently, some large ORs reported may be an overestimation of the genuine co-morbidity.

To address the limitations of previous studies, the aim of the current chapter is to establish whether CUD and MDD are co-morbid in a cohort of 3824 Australian twins and their non-twin siblings (see Section 2.1) and to what extent this co-morbidity is influenced by covariates. The influence of covariates will be demonstrated by the extent to which the overall association between MDD and CUD diminishes when covariates are accounted for. Additionally, the influence of individual covariates will be analysed and interpreted. Data analysed in this thesis has been collected for the purpose of examining cannabis-related research questions (Lynskey et al., 2012). Consequently, it is enriched for known covariates and is well-positioned to provide

an estimate of co-morbidity which is less likely to be confounded by unexamined covariates.

The main aim of the current chapter is to report epidemiological analyses, which address two main questions:

1. Is there a significant association between CUD and MDD in the twin sample examined in this thesis?

2. What role do covariates play in this relationship?

    a. Which covariates influence CUD and MDD uniquely, which influence both disorders?

For the first research question, a logistic regression model will be used to estimate the odds ratio between binary measures of DSM-5 CUD and MDD, controlling for age and sex only. Addressing the second aim will involve fitting a series of logistic regression analyses, again focused on CUD and MDD, but in which covariates significantly associated with cannabis involvement and/or depression in previous literature will be controlled for. These covariates include age, sex, family SES, other substance dependence (nicotine, alcohol and illicit drugs), conduct disorder, panic disorder, social phobia, childhood sexual abuse, death of a parent, having been raised by both parents, the excessive drinking of parents, disagreements with parents, parental problems, and peer drug use. The adjusted odds ratio (aOR) will be an indicator of the extent to which the relationship between CUD and MDD changes when known covariates are taken into account.

Testing whether the association between CUD and MDD is significant in this cohort is not only a contribution to previous literature, but also plays a central role for subsequent twin model analyses in this thesis. Twin model analyses decompose the covariance between two phenotypes into genetic and environmental factors. Therefore, establishing whether there is a significant covariance between two phenotypes in the first place is a prerequisite for such analyses.

4.2 Methods

4.2.1 Sample

This chapter outlines analyses on the Australian twin cohort which was comprehensively described in Section 2.1. The analysis sample consisted of 3824 MZ and DZ twins, as well as their non-twin siblings, who were born between 1972 and 1979.

4.2.2 Measures

4.2.2.1 Main variables

Details on the definition and coding of the main variables of interest – CUD and MDD – can be found in Chapter 2. Briefly, MDD and CUD were coded as binary variables and according to DSM-V criteria as far as the information included in the SSAGA-OZ interview allowed.

4.2.2.2 Covariates

1.1.1.1.4 Choice of covariates

To assemble a list of covariates for inclusion in the epidemiological analyses, studies reviewed in Chapter 1 were examined, in particular those which explored a variety of covariates in their analyses and reported adjusted and unadjusted ORs (see Table 2). These ORs indicate to what extent the included covariates may mediate the relationship between CUD and MDD. Due to the heterogeneity of covariates tested in the literature, they were grouped into categories.

The following categories were identified:

- Demographic factors
- Family or parental factors
- Childhood behavioural problems (e.g. conduct problems)
- Childhood trauma

- Other psychiatric/psychological problems
- Other drug use/use disorders
- Romantic relationships
- Peer factors
- Neighbourhood factors
- Educational/work factors
- Adolescent conduct problems and drug use
- Personality/intelligence factors
- General health and life satisfaction

These categories were then mapped as closely as possible onto the covariate measures available in the current dataset. Only those items with information on the age of onset or items which likely occurred before the onset of CUD and MDD, i.e. in childhood or adolescence, were further considered. This was done to avoid over controlling for variables which may have been caused by rather than the cause of CUD or MDD, and which thus may lead to an underestimation of the true co-morbidity.

The selected covariates are summarised in Table 13. Not all categories identified in the literature could be matched with items present in the interview. This was either because items were unavailable, age of onset was not measured, or the variable was not likely to have occurred before the CUD and MDD ages of onset.

For instance, none of the items assessed under the ‘romantic relationships’ category, such as marriage, divorce and separation were likely to, on average, occur before the age of onset of CUD. Therefore, no covariates pertaining to romantic relationships were assessed. Other examples are general health and work (e.g. employment status) factors. While questions on both were available in the interview and matched the categories identified from the literature, none asked about ages of onset in this dataset. If changes in physical health or employment status were consequences, rather than causes of depressive symptoms or cannabis involvement, including these in the model would have overcontrolled for covariates and possibly masked the relationship between CUD and MDD.

In addition to the variables presented in Table 13, ADHD and Agoraphobia were also initially considered. However, they were excluded, because the number of ADHD cases was low (N = 24) and Agoraphobia showed high collinearity with Panic Disorder.

Similar to many other studies investigating the relationship between cannabis involvement and depression (e.g. Brook, Zhang, & Brook, 2011; Degenhardt et al., 2013; Marmorstein & Iacono, 2011), most covariates were measured as either binary or ordinal variables.

Table 13. Covariates examined in the epidemiological analyses.

Variable	Question	Age of Onset	Type	Sample Characteristics
Demographics				
Age	'How old are you now?'	-	Continuous	Min = 21 Max= 46
Gender	'Male/Female'	-	Binary	Female = 2435 Male = 1389
Family SES	'Compared with the average family in your community when you were 6 to 13, was your family financially better off, about average, or worse off during most of that time?'	6–13	Ordinal	Better off = 618 Average = 2687 Worse off = 511
Family/Parental				
Raised by both parents until age 16?	'Were you raised by both of your biological parents until age 16?'	<16	Binary	Yes = 3149 No = 674
Disagreements with mother	'How often did you have an unpleasant disagreement or conflict with your mother (mother figure) between the ages of 6 to 13?'	6–13	Ordinal	Often = 258 Sometimes = 1155 Rarely = 2005 Never = 387
Disagreements with father	'How often did you have an unpleasant disagreement or conflict with your father (father figure) between the ages of 6 to 13?'	6–13	Ordinal	Often = 405 Sometimes = 860 Rarely = 1478 Never = 861
Parental fighting	'Between the ages of 6 to 13, how often did your parents (parent figures) fight or argue in front of you?	6–13	Ordinal	Never = 862 Rarely = 1478 Sometimes = 860 Often = 405

Variable		Question	Age of Onset	Type	Sample Characteristics
Parental conflict/tension		'How much conflict and tension was there between your parents (parent figures) when you were 6 to 13?'	6–13	Ordinal	A lot = 385 Some = 641 A little = 1384 None = 1192
Mother excessive drinking		'Do you now think your mother (mother figure) drank too much when you were 6 to 13?'	6–13	Binary	Yes = 138 No = 2618
Father excessive drinking		'Do you now think your father (father figure) drank too much when you were 6 to 13?'	6–13	Binary	Yes = 2581 No = 623
Mother strictness		'In your opinion, when you were 6 to 13, was your mother (mother figure) more strict than most mothers?'	6–13	Binary	No = 2577 Yes = 1215
Father strictness		'In your opinion, when you were 6 to 13, was your father (father figure) more strict than most fathers?'	6–13	Binary	No = 2337 Yes = 1273
Inconsistent parenting	Mother	'Was your mother (mother figure)/ father (father figure) pretty consistent about the rules, or was she more inconsistent, where she would make you follow a rule one day and forget about it the next?'	6–13	Binary	No = 3456 Yes = 339
	Father		6–13	Binary	No = 3113 Yes = 496
Trauma					
Death of father		'Is your biological father still alive?'	'How old were you when he died?'	Binary	**Before CUD or MDD onset** No = 3571 Yes = 195 **Lifetime** No = 3456 Yes = 322

Variable	Question	Age of Onset	Type	Sample Characteristics
Death of mother	'Is your biological mother still alive?'	'How old were you when she died?'	Binary	**Before CUD or MDD onset** No = 3703 Yes = 98 **Lifetime** No = 3639 Yes = 174
Childhood Sexual Abuse	'Before age 18, were you ever forced into sexual intercourse or any other form of sexual activity?'	'How old were you the first time you were forced into sexual activity?'	Binary	**Before CUD or MDD onset** No = 3476 Yes = 292 **Lifetime** No = 3457 Yes = 324
Neighbourhood Factors				
School Safety	'How safe was the school that you last attended?'	High school	Ordinal	Very unsafe = 69 Somewhat unsafe = 188 Pretty safe = 1337 Very safe = 2224
Peer factors				
Peer illegal drug use	'In your opinion, how many of the students who were in the same grade as you at the last high school you attended ever used illegal drugs such as marijuana while of high school age?	High school	Ordinal	None = 185 Just one or two = 444 Just a few = 1545 A quarter = 948 Half = 413 Three-quarters = 195 Almost all = 49

Variable	Question	Age of Onset	Type	Sample Characteristics
Peer alcohol use	'In your opinion, how many of the students who were in the same grade as you ever used alcohol while of high school age?'	High school	Ordinal	None = 46 Just one or two = 129 Just a few = 569 A quarter = 733 Half = 857 Three-quarters = 839 Almost all = 638
Peer cigarette use	'How many of the students who were in the same grade as you ever smoked cigarettes while of high school age?	High school	Ordinal	None = 33 Just one or two = 195 Just a few = 1311 A quarter = 1274 Half = 693 Three-quarters = 241 Almost all = 62
Other drug use				
Nicotine Dependence	All individuals who reported smoking at least 100 times in their lifetime, were assessed on DSM-IV criteria of nicotine dependence. The binary response to the clustering question (3+ symptoms in a 12-month period) was used as the indicator of nicotine dependence in the analyses.	Onset of *any* nicotine dependence symptom in DSM-IV questionnaire (SSAGA-OZ)	Binary	**Before CUD or MDD onset** No = 3059 Yes = 749 **Lifetime** No = 2845 Yes = 976
Alcohol Dependence	All individuals who reported drinking at least five drinks in a single day at least once, were assessed on DSM-IV criteria of alcohol dependence. Those reporting three or more symptoms occurring within a 12-month period were classified as meeting criteria for DSM-IV alcohol dependence.	Onset of *any* alcohol dependence symptom in DSM-IV questionnaire (SSAGA-OZ)	Binary	**Before CUD or MDD onset** No = 3091 Yes = 707 **Lifetime** No = 2862 Yes = 949

Variable	Question	Age of Onset	Type	Sample Characteristics
Other illicit drug dependence	All individuals who reported using any illicit drug (cocaine, stimulants, opiates, sedatives, hallucinogens, PCP, solvents, inhalants) at least 10 times over their lifetime, were assessed on DSM-IV criteria of illicit drug dependence. Those reporting 3+ symptoms in a 12-month period were classified as meeting DSM-IC criteria for illicit drug dependence.	Onset of *any* illicit drug dependence symptom in DSM-IV questionnaire (SSAGA-OZ)	Binary	**Before CUD or MDD onset** No = 3717 Yes = 68 **Lifetime** No = 3618 Yes = 180
Other psychiatric problems/childhood behavioural problems				
Social Phobia	Social phobia was coded as present if participant met DSM-IV criteria (SSAGA-OZ) of social phobia (except for criterion F, which was not measured), and none of the exclusion criteria.	'How old were you when you first began to: Avoid this situation or do it feeling very uncomfortable? […] Feel extremely nervous or panicky right away in any of these situations?'	Binary	**Before CUD or MDD onset** No = 3266 Yes = 505 **Lifetime** No = 3198 Yes = 586
Panic Disorder	Panic disorder was coded as present if participants met DSM-IV criteria (SSAGA-OZ) of panic disorder. Exclusion criteria were applied, except that not sufficient information was available to exclude cases that would be better accounted for by other psychiatric disorders.	'How old were you the first time you had one of these sudden attacks/sudden periods of physical discomfort along with four (or more) problems from this list?'	Binary	**Before CUD or MDD onset** No = 3707 Yes = 64 **Lifetime** No = 3642 Yes = 142
Conduct Disorder	Individuals were assessed on DSM-IV Conduct Disorder criteria (SSAGA-OZ) before age 18	'How old were you the first time you did at least three of these things within the same 12-month period?'	Binary	**Before CUD or MDD onset** No = 3502 Yes = 264 **Lifetime** No = 3459 Yes = 325

4.2.3 Statistical analyses

4.2.3.1 Logistic regression analyses

To assess the co-morbidity between MDD and CUD, and the influence of covariates on this relationship, several logistic regression models were conducted (see Chapter 2 for a detailed description). The first model examined the OR between CUD and MDD and controlled for sex and age only. The second is a full model, controlling for all measured covariates. The third model only includes predictors at $\alpha < 0.1$ measured by backward selection. As outlined in Chapter 2, reducing the model also reduces the effect of multicollinearity, leading to more accurate estimates of standard errors. The second and third model were fit with CUD and MDD as outcome variables, to a) remain agnostic with regard to the direction of causation, and b) assess shared and unique influences of covariates. Statistical analyses were conducted in STATA (StataCorp, 2015) and the Huber-White correction was applied to account for family clustering.

4.2.3.2 Controlling for age of onset

As mentioned above, the regression models only aimed to control for factors which were likely to contribute to the onset of CUD and MDD and thereby explain their co-occurrence. Since the average age of onset for CUD was 19.9 and for MDD 23.8, only childhood and adolescent factors or those variables which had a measured age of onset were controlled for.

Variables which included age of onset information were only coded as 'present' if they occurred before the age of onset of CUD *and* MDD symptoms. As an example, if a participant met the criteria for nicotine dependence over their lifetime, they were coded as 1 ('present') only if the age of onset of their nicotine dependence was before the age of onset of their CUD *and* MDD. Cases where CUD or MDD criteria were met, but either age of onset was missing, were deleted from the analyses.

Factors which did not have a defined age of onset were judged as to their likelihood of occurring before or after any onset of a CUD or MDD episode. Where a covariate

spanned a wider range of time, the variable was included to provide a conservative estimate of the adjusted OR between CUD and MDD. For instance, some individuals may have had a CUD onset as early as high school, which would overlap with the peer drug intake measures. However, observing peer drug use most likely happened before any CUD symptoms and certainly before the average age of CUD onset.

The age of onset variable chosen for MDD was the onset of the most severe period of depression that did not start within two months of bereavement. It was also the age of onset variable that had the least missing data among individuals who met criteria for MDD (missing N = 1). The mean age of onset for this most severe period of depression was 23.82 years. A question about the age of onset alongside each MDD symptom check list would have been the preferred way to measure MDD onset, but was unavailable in the dataset.

For CUD, the age of onset was available for each dependence criterion measured in the symptom checklist. However, no age of onset was available for abuse criteria. Since there were 12 cases where the age of onset was unavailable, those may have been CUD cases that only met the abuse criteria.

4.3 Results

4.3.1 The association between CUD and MDD adjusted for age and sex

The results of a logistic regression with MDD as the outcome variable and CUD as the predictor are shown in Table 14. Age and sex are controlled for. The same OR would be estimated with CUD as the outcome and MDD as the predictor.

Table 14. Prevalence of and odds ratios (95% confidence intervals) between MDD and CUD, controlled for age and sex.

Odds Ratio	2.23 (1.84 – 2.70)	
	No MDD	MDD
No CUD	88.19%	77.04%
CUD	11.81%	22.96%

4.3.2 Covariates associated with CUD or MDD

After examining odds ratios between the variables of interest and all variables in the dataset that matched the categories extracted from previous studies, only those variables with a significant relationship with either variable of interest were kept for the logistic regression analyses. Table 15 shows the ORs between the covariates and both MDD and CUD. Note that not all variables presented in Table 13 are presented here. Firstly, only factors significantly associated with at least one variable of interest are shown. Secondly, several related variables were combined due to high and significant polychoric correlations between them, or to increase cell counts: a) ‘peer illegal drug’, ‘alcohol’ and ‘cigarette use’ were combined into a binary variable indicating ‘any peer drug use’; b) ‘death of father’ and ‘death of mother’ were combined into a binary ‘death of parent’ variable; c) ‘disagreements with mother’ and ‘disagreements with father’ were also combined into ‘disagreements with parents’, with the same response options of ‘never’, ‘rarely’, ‘sometimes’ and ‘often’ as the original variables; d) ‘inconsistent parenting’ and ‘strict parenting’ were combined from binary variables specific to each parent, and the same response categories were kept. A ‘parental problems’ count variable was created from four binary variables: ‘parental fighting’, ‘parental conflict’, ‘inconsistent parenting’ and ‘strict parenting’.

Table 15. Unadjusted odds ratios (95% confidence intervals) between CUD, MDD and selected measures of demographic variables, family factors, childhood trauma, neighbourhood factors, peer factors, other drug use and other psychiatric problems.

Variable	OR MDD (95% CI)	OR CUD (95% CI)
Demographics		
Age	1.01 (0.99–1.04)	0.94 (0.91–0.98)
Gender (male = 0, female = 1)	1.93 (1.62–2.29)	0.43 (0.36–0.53)
Family SES	0.94 (0.76–1.17)[1] 1.73 (1.32–2.27)[2]	0.81 (0.63–1.05)[1] 0.86 (0.61–1.21)[2]
Family/Parental		
Raised by both parents until age 16?	0.67 (0.55–0.81)	0.61 (0.48–0.77)
Disagreements with parents	1.27 (0.90–1.79)[3] 2.00 (1.42–2.82)[4] 4.27 (2.95–6.17)[5]	1.53 (0.93–2.51)[3] 2.47 (1.51–4.04)[4] 3.79 (2.27–6.34)[5]
Parental problems[3] (parental fighting, conflict, inconsistent and strict parenting)	1.95 (1.58–2.41)[6] 3.61 (2.89–4.50)[7] 5.13 (3.81–6.90)[8]	1.44 (1.12–1.86)[6] 1.79 (1.37–2.32)[7] 2.34 (1.65–3.33)[8]
Parents excessive drinking	1.64 (1.36–1.98)	1.20 (0.95–1.52)
Trauma		
Death of parent	0.85 (0.62–1.17)	0.63 (0.40–0.98)
Childhood Sexual Abuse	3.01 (2.35–3.85)	1.97 (1.46–2.65)
Neighbourhood Factors		
School Safety[4]	1.19 (1.02–1.39)[9] 1.67 (1.26–2.20)[10]	1.46 (1.20–1.77)[9] 2.02 (1.45–2.80)[10]
Peer factors		
Any peer drug use	1.15 (0.92–1.44)	2.59 (1.80–3.73)
Other drug use		
Nicotine Dependence	1.56 (1.31–1.87)	2.94 (2.42–3.57)
Alcohol Dependence	1.24 (1.03–1.49)	1.34 (1.07–1.67)
Other illicit drug dependence	2.01 (1.21–3.35)	3.55 (2.13–5.90)
Other psychiatric problems		
Social Phobia	3.68 (3.03–4.47)	1.23 (0.94–1.61)
Panic Disorder	5.42 (3.19–9.23)	1.85 (0.99–3.46)
Conduct Disorder	2.22 (1.70–2.90)	6.22 (4.77–8.11)

Note. 1) "Better off" vs "Average"; 2) "Better off" vs "Worse off"; 3) "Never" vs "Rarely"; 4) "Never" vs "Sometimes"; 5) "Never" vs "Often"; 6-8) 0 vs 1-3; 9) "Very safe" vs "Pretty safe", 10) "Very safe" vs "Somewhat/very unsafe"

4.3.2.1 Covariates associated with CUD and MDD

Most covariates were independently associated with both phenotypes. Between the ages of 6 and 13, individuals with MDD and participants with CUD were more likely to have been raised by a single parent, to have had disagreements with both parents, and to have experienced higher levels of parental problems, which include parental fighting, conflict or tension, and above average strict or inconsistent parenting. Any drug dependence – nicotine, alcohol and illicit drugs – was more likely to occur in both groups, as was conduct disorder, childhood sexual abuse, and attending a school that was not self-rated as 'very safe'. Finally, there are significant gender differences for both CUD and MDD. While men were more likely to experience CUD, women were more likely to meet criteria for MDD.

4.3.2.2 Covariates associated only with MDD

In addition to the above factors, individuals with MDD were also more likely to report being from a family that was 'worse off' financially and more likely to believe that their parents drank excessively when they were 6 to 13 years old. They also had higher odds of meeting criteria for social phobia and panic disorder.

4.3.2.3 Covariates associated only with CUD

Individuals with CUD were younger, and were more likely to report that their peers consumed illegal drugs, alcohol or cigarettes. Finally, participants who lost either parent before the onset of MDD and CUD were also significantly less likely to report CUD.

4.3.3 Associations between CUD and MDD adjusted for covariates

Given that multiple variables were associated with both phenotypes of interest, the aim of the following analyses is to assess, firstly, whether the co-morbidity between MDD and CUD stays statistically significant once covariates are considered. Secondly, the analyses examine which covariates play a particularly important role when all variables are assessed simultaneously.

4.3.3.1 Predictors of MDD

4.3.3.2 Full Model

The full model with MDD as an outcome shows that the OR between CUD and MDD remains statistically significant even after controlling for all covariates (see Table 16). Each level of an ordinal variable (e.g. parental problems) was treated as a separate variable by the model. Individuals with CUD have almost twice the odds of meeting the criteria for MDD (OR = 1.92 (95% CI = 1.52–2.42)). The Wald statistic indicates that the predictors significantly contribute to the model (Wald $\chi^2(24) = 429.32$, $p = <.001$), and they explain 12% of the variance of MDD (*pseudo-*$R^2 = .12$). The mean VIF is 3.64. This is mainly driven by the large VIF for age (VIF = 27.01), which is highly correlated with many other variables in the model. No other variable had a VIF above 10. There are no clear cut-offs for VIFs which indicate problematic multicollinearity, but VIFs below 10 are often cited as acceptable (Field, Miles, & Field, 2012).

Table 16. Full multivariable regression model with odds ratios (95% confidence intervals) between selected predictors and Major Depressive Disorder (MDD).

MDD	Odds Ratio (95% CI)	Robust SE	*p*-value*
Cannabis Use Disorder	1.92 (1.52–2.42)	0.12	<.001
Age	1.02 (0.99–1.05)	0.01	.215
Sex	1.92 (1.58–2.34)	0.10	<.001
Family SES			
Average	1.01 (0.81–1.27)	0.12	.925
Worse off	1.45 (1.08–1.95)	0.15	.014
Nicotine Dependence	1.09 (0.89–1.35)	0.11	.402
Alcohol Dependence	1.17 (0.94–1.45)	0.11	.152
Illicit Drug Dependence	1.15 (0.64–2.08)	0.30	.645
Conduct Disorder	1.47 (1.06–2.04)	0.17	.020
Panic Disorder	3.03 (1.66–5.51)	0.31	<.001
Social Phobia	2.86 (2.31–3.53)	0.11	<.001
Raised by both parents	1.01 (0.80–1.26)	0.12	.966
Excessive drinking parents	0.99 (0.80–1.23)	0.11	.945
Childhood Sexual Abuse	1.70 (1.27–2.27)	0.15	<.001
Death of parent	0.76 (0.54–1.08)	0.18	.131
Disagreement with parents			
Rarely	1.01 (0.71–1.42)	0.18	.967
Sometimes	1.24 (0.88–1.77)	0.18	.221
Often	1.85 (1.24–2.76)	0.20	.002
Parental Problems			
1	1.51 (1.20–1.90)	0.12	<.001
2	2.19 (1.70–2.82)	0.13	<.001
3	2.60 (1.84–3.66)	0.18	<.001
School Safety			
Pretty safe	1.03 (0.86–1.22)	0.09	.776
Very/somewhat unsafe	1.28 (0.92–1.77)	0.17	.138
Peer drug use	1.02 (0.81–1.30)	0.12	.828

Note. pseudo-R^2 = .12, Wald $\chi^2(24) = 429.32$; *Bonferroni adjusted

4.3.3.3 Backward regression

The results of a backward regression with an inclusion *p*-value of .100 are presented in Table 17. The number of variables in the model was reduced from 24 to 12 and the Wald χ2 was statistically significant at p <.001. Approximately 12% of the variation in MDD was explained by the predictors in the model. The average VIF within this model was 1.45, with the maximum VIF being 2.16 for sex. This indicates

that standard errors are, on average, 1.45 times larger than if the predictors were not correlated.

All predictors included in this model were significant at a Bonferroni-corrected theshold of 0.05. Bearing in mind the full list of possible predictors, demographic, family problems, and other psychopathologies were particularly strong predictors of MDD, as was childhood sexual abuse.

Table 17. Odds ratios (95% confidence intervals) for significant predictors of Major Depressive Disorder (MDD) after backward regression.

MDD	Odds Ratio (95% CI)	Robust SE	*p*-value*
Cannabis Use Disorder	1.96 (1.57–2.45)	0.12	<.001
Sex	1.86 (1.53–2.26)	0.10	<.001
Family SES			
Worse off	1.44 (1.15–1.80)	0.11	<.001
Conduct Disorder	1.55 (1.13- 2.12)	0.16	.006
Panic Disorder	3.12 (1.71–5.67)	0.30	<.001
Social Phobia	2.87 (2.32–3.54)	0.11	<.001
Disagreement with parents			
Sometimes	1.26 (1.05–1.52)	0.10	.015
Often	1.88 (1.45–2.44)	0.13	<.001
Parental Problems			
1	1.51 (1.20–1.90)	0.12	<.001
2	2.22 (1.73–2.85)	0.13	<.001
3	2.62 (1.88–3.67)	0.17	<.001
Childhood Sexual Abuse	1.73 (1.30–2.30)	0.15	<.001

Note. pseudo-R^2 = .12, Wald $\chi^2(12) = 415.61$; *Bonferroni adjusted

4.3.3.4 Predictors of CUD

4.3.3.5 Full Model

The full model with CUD as an outcome (see Table 18) also shows an approximately twofold increase in the odds of also meeting criteria for MDD (OR = 1.90, 95% CI = 1.51 – 2.40). It explains approximately 15% of the variation in CUD and predictors significantly contribute to the model (Wald $\chi^2(24) = 380.71$, p = <.001). The average VIF was 3.65, mainly influenced by the large VIF for age (VIF = 27.00). No other variable exceeded a VIF of 10.

Table 18. Full multivariable regression model with odds ratios (95% confidence intervals) between selected predictors and Cannabis Use Disorder (CUD).

CUD	Odds Ratio (95% CI)	Robust SE	*p*-value*
Major Depressive Disorder	1.90 (1.51–2.40)	0.12	<.001
Age	0.94 (0.91–0.98)	0.02	.001
Sex	0.37 (0.29–0.46)	0.12	<.001
Family SES			
Average	0.99 (0.74–1.33)	0.15	.969
Worse off	0.69 (0.47–1.03)	0.20	.069
Nicotine Dependence	2.47 (1.97–3.10)	0.12	<.001
Alcohol Dependence	0.72 (0.56–0.94)	0.13	.016
Illicit Drug Dependence	1.91 (1.01–3.62)	0.33	.046
Conduct Disorder	3.36 (2.47–4.58)	0.16	<.001
Panic Disorder	0.97 (0.45–2.08)	0.39	.930
Social Phobia	0.90 (0.66–1.24)	0.16	.524
Raised by both parents	0.64 (0.48–0.84)	0.14	.002
Excessive drinking parents	0.97 (0.73–1.29)	0.14	.852
Childhood Sexual Abuse	1.78 (1.24–2.57)	0.19	.002
Death of parent	0.57 (0.36–0.92)	0.24	.022
Disagreement with parents			
Rarely	1.34 (0.81–2.22)	0.26	.252
Sometimes	1.84 (1.11–3.07)	0.26	.018
Often	2.09 (1.19–3.69)	0.29	.011
Parental Problems			
1	1.25 (0.95–1.66)	0.14	.112
2	1.21 (0.89–1.65)	0.16	.227
3	1.30 (0.82–2.04)	0.23	.258
School Safety			
Pretty safe	1.09 (0.87–1.35)	0.11	.453
Very/somewhat unsafe	1.19 (0.81–1.75)	0.20	.379
Peer drug use	2.18 (1.47–3.23)	0.20	<.001

Note. *pseudo-*R^2= .15, Wald $\chi 2(24) = 380.71$; *Bonferroni adjusted

4.3.3.6 Significant Predictors of Cannabis Use Disorder

The reduced model after backward selection included 14 predictors of CUD. It explained around 15% of the variance in CUD and was an adequate model (Wald $\chi^2(12) = 370.06$, p = <.001). The variance in CUD explained by the reduced model

was not smaller than that in the full model. The mean VIF for this model was 3.13, again largely due to the large VIF of age (VIF = 14.67). Somewhat higher VIFs were also found for peer drug use (VIF = 6.86) and being raised by both parents (VIF = 6.39). Standard errors for these variables are likely to be inflated due to correlations with other predictors.

All predictors in the reduced model are significant at a Bonferroni-corrected threshold of 0.05. Demographic, family and other drug use variables are particularly strong predictors of CUD, as are conduct disorder and childhood sexual abuse. Individuals with CUD are also less likely to report the death of a parent.

Table 19. Odds ratios (95% confidence intervals) for significant predictors of Cannabis Use Disorder (CUD) after backward regression.

CUD	Odds Ratio (95% CI)	Robust SE	*p*-value*
Major Depressive Disorder	1.92 (1.53–2.41)	0.12	<.001
Age	0.94 (0.91–0.97)	0.02	.001
Sex	0.37 (0.29–0.46)	0.12	<.001
Family SES			
Worse off	0.70 (0.51–0.96)	0.16	.025
Nicotine Dependence	2.50 (2.00–3.12)	0.11	<.001
Alcohol Dependence	0.73 (0.56–0.94)	0.13	.016
Illicit Drug Dependence	1.89 (1.01–3.56)	0.32	.048
Conduct Disorder	3.44 (2.52–4.69)	0.16	<.001
Raised by both parents	0.62 (0.47–0.82)	0.14	.001
Disagreement with parents			
Sometimes	1.48 (1.19–1.86)	0.11	.001
Often	1.72 (1.24–2.38)	0.17	.001
Peer drug use	2.19 (1.48–3.24)	0.20	<.001
Death of parent	0.57 (0.36–0.92)	0.24	.021
Childhood Sexual Abuse	1.79 (1.24–2.60)	0.19	.002

Note. pseudo-R^2= .15, Wald $\chi^2(12) = 370.06$; *Bonferroni adjusted

4.4 Discussion

4.4.1 Association between CUD and MDD

This chapter aimed to establish the co-morbidity between MDD and CUD in a sample of 3824 Australian twins and siblings, and to elucidate the role covariates play in this relationship. There are two principal findings that emerged from these analyses:

1. When controlling for age and sex only, meeting criteria for one of the disorders increased the odds of the other by 2.23 times.

2. These odds were not significantly diminished, indicated by overlapping confidence intervals of the adjusted and unadjusted associations, when a multitude of covariates predicting either phenotype were controlled for. After backward selection, the OR between CUD and MDD was 1.96 (with MDD as outcome) and 1.92 (with CUD as outcome).

Overall, these results suggest that the CUD and MDD remain robustly associated in a large sample, despite controlling for an extensive range of covariates identified from previous literature.

4.4.1.1 Shared covariates

Although none of the covariates significantly attenuated the association between MDD and CUD, the somewhat attenuated OR may be due to several common covariates influencing the phenotypes of interest. After backward selection, disagreements with parents, having experienced childhood sexual abuse and meeting the criteria for conduct disorder significantly increased the likelihood of reporting both MDD and CUD. Sex and family SES also influenced both phenotypes significantly, however in opposite directions: males were more likely to report CUD, while females were more likely to report MDD. Additionally, individuals from families which were ‘worse off’ financially had significantly higher odds of reporting MDD, but were significantly less likely to meet criteria for CUD.

Controlling for these covariates may be necessary to obtain an accurate association between MDD and CUD, and the lack of controlling for them may explain the differences in ORs reported in previous cross-sectional and longitudinal studies.

4.4.1.2 Unique covariates

The analyses also highlighted a number of variables which uniquely influenced each phenotype. CUD was uniquely influenced by drug-related factors: nicotine, alcohol and illicit drug dependence, as well as peer drug use. Additionally, individuals with CUD were younger and less likely to be raised by both parents. Finally, CUD was less likely among individuals who had experienced the death of a parent. Meanwhile, individuals with MDD experienced significantly higher levels of parental problems, and were more likely to meet the criteria for social phobia and panic disorder.

4.4.2 Integration of findings with previous literature and implications

These results, tentatively, add evidence to the idea of a unique aetiological pathway between CUD and MDD, which cannot be fully explained by covariates. This conclusion is further supported by results from previous research that has employed a discordant twin approach in MZ twins, since all genetic (A and D) and shared environmental (C) factors are inherently controlled for. A discordant twin study of 1874 MZ males found that twins with DSM-III-TR Major Depressive Disorder had 4.4 (95% CI = 2.7–7.0) times higher odds of reporting DSM-III-TR cannabis abuse/dependence than their co-twins without MDD (Lin et al., 1996). A recent discordant twin study, including the current sample in addition to two other Australian twin cohorts, has found a significantly increased risk (OR = 1.72, 95% CI = 1.05–2.82) of MDD in male and female twins who engaged in frequent cannabis use, compared to their co-twins who did not (Agrawal et al., 2017). A previous study by Lynskey et al. (2004) has not found a significant association between cannabis dependence and MDD in MZ twins. However, the analysed cohort in this previous study was also included in the larger discordant twin study, and the authors suggest that the absence of a significant OR may have been due to sample size issues. This is likely given that analyses can only be conducted on twin pairs discordant for both phenotypes, which will be a small sub-sample of twins.

As mentioned in Chapter 1, placing the current findings in the context of previous cross-sectional and longitudinal research is complicated by the heterogeneity of that literature. Studies have measured cannabis involvement and depression in various ways, controlled for different confounding variables and few studies having specifically examined CUD and MDD. Additionally, samples were taken from varying regions, ethnic makeups and age ranges. Variations in results therefore need to be interpreted bearing in mind these differences. These caveats aside, the current study results are compatible with several studies which were reviewed in Chapter 1 and have examined the association between clinical definitions of cannabis involvement and depression.

Several cross-sectional and longitudinal studies found a significant association between clinical levels of cannabis involvement and depression which did not diminish with adjustment for covariates. Martins and Gorelick (2011) found that individuals with a lifetime history of a DSM-IV mood disorder (MDD, dysthymia, mania, hypomania) were 3.9 (aOR, 95% CI = 2.8–5.3, N = 43093) times more likely to meet criteria for lifetime DSM-IV cannabis abuse/dependence. Pacek et al. (2013) examined the bi-directional longitudinal association between MDD and CUD, where participants meeting the criteria for baseline CUD were significantly more likely to develop MDD in the following three years (OR = 2.02, 95% CI = 1.35–3.04, N = 43093). This OR was reduced to 1.78 (95% CI = 1.17–2.71) when adjusted for demographic variables and family history of depression, but remained significant. Participants with baseline MDD showed similarly higher odds of developing CUD within three years (OR = 2.01, 95% CI = 1.09–3.68; aOR = 2.28, 95% CI = 1.28–4.05). Bovasso (2001) reported a fourfold (aOR = 4.00, 95% CI = 1.23–12.97) increase in the odds of developing DSM-III-R depressive symptoms in individuals who reported DSM-III cannabis abuse approximately 15 years prior. Finally, Marmorstein and Iacono, (2011) reported an aOR of 2.54 (95% CI = 1.40–4.60) between DSM-III-R-based CUD at age 17 and DSM-III-R MDD within a 7-year follow-up period.

However, several studies did not report significant co-morbidity in the first place, or found that it became non-significant after adjustment. To name some examples among the studies which have been reviewed in Chapter 1, a longitudinal study on

an adolescent sample did not find a significant relationship between DSM-IV Cannabis Abuse or Dependence and Major Depressive Disorder (Harder et al., 2008), and Degenhardt et al. (2013) found no significant association between DSM-IV Cannabis Dependence at ages 20 or 24 with MDD at age 29. Degenhardt et al. (2001) report an unadjusted OR of 2.88 (95% CI = 1.61–5.17, N = 10641) between DSM-IV Cannabis Abuse and DSM-IV Affective disorders (MDD, dysthymia, bipolar I and II), and of 2.85 (95% CI = 1.86–4.35) between Cannabis Dependence and Affective disorders, but these associations became non-significant when demographics and other drug use were adjusted for in the model.

Overall, the current study results fall within the range of results found in previous twin, longitudinal and cross-sectional studies, and are compatible with the interpretation that there is a unique link between CUD and MDD which cannot fully be accounted for by covariates.

Additionally, a cluster of childhood factors has been identified in the analyses, which could be the target of early interventions for the co-morbidity between MDD and CUD. Children who met criteria for conduct disorder, reported childhood sexual abuse and disagreed with their parents sometimes or often were more likely to meet the criteria for both disorders. These factors have been linked to both depression and cannabis involvement in the previous literature (see Table 2; Fergusson & Horwood, 1997; Hornor, 2009). Furthermore, both conduct disorder (Furlong et al., 2013; Ogden & Hagen, 2008) and childhood sexual abuse (Fergusson, Lynskey, & Horwood, 1996; Putnam, 2018) may be linked to problematic parental behaviour. A recent Cochrane review has concluded that parent training has been shown to significantly reduce conduct problems in children (Furlong et al., 2013). Although parent-centred interventions to help prevent childhood sexual abuse have not yet been widely implemented, they are a promising area for research and practice (Mendelson & Letourneau, 2015). Providing parent training is not only beneficial for children, but also improves parental mental health (Furlong et al., 2013). Consequently, previous literature and findings from this thesis add evidence to support parenting interventions as a preventative measure for later mental health problems in general, and depression and cannabis-related problems in particular.

4.4.3 Strengths and limitations

Data collection for the cohort of Australian twins was tailored around cannabis and cannabis-related variables, which allowed a comprehensive investigation of covariates in the current analyses. Although the data were retrospective (this and other strengths and limitations which apply to all analyses in this thesis will be discussed in Section 7.3), the age of onset was collected for many relevant covariates, allowing to mitigate the risk of controlling for variables which may have been the result of CUD or MDD.

Nevertheless, epidemiological analyses on twin samples may need to be interpreted with some caution. Twins may differ from the general population in a number of ways. One example is an increased risk of obstetric complications (Kendler, Martin, Heath, & Eaves, 1995). However, studies on various measures of psychopathology, including depression, have not found significant differences in the rates of the examined disorders between twins and non-twins (Kendler, Martin, et al., 1995; Kendler, Pedersen, Farahmand, & Persson, 1996; Rutter & Redshaw, 1991). Additionally, the association reported in the current study is compatible with findings in non-twin populations. Therefore, the external validity of the conclusions in this chapter is not likely to be limited due to analysing twin data.

An additional limitation in the analyses has resulted from the difficulty of matching variables in the dataset with some categories of variables – identified from previous literature – which were found to be related to cannabis involvement and depression. Matches with ages of onset before both CUD and MDD have been found for the majority of categories, but not all. Specifically, no variable related to ‘romantic relationships’, ‘educational/work factors’, ‘personality/intelligence’ or ‘general health and life satisfaction’ were included in the analyses. Although the dataset contained measures related to all these categories, divorces or having occupational difficulties, to mention two examples, were on average more likely to occur later than symptoms of CUD or MDD. The decision to exclude these additional variables may explain why the regression models, whether predicting CUD or MDD, explained a small portion of variance in the phenotype. This indicates that a more extensive inclusion of covariates would be important. However, the current sample required a trade-off

between including covariates comprehensively and including covariates that were likely to have occurred before the onset of either phenotype of interest. A more accurate control for covariates would have been possible with a longitudinal sample.

4.4.4 Conclusions and further analyses

Results from the presented epidemiological analyses suggest that MDD and CUD are robustly associated in this cohort. More specifically, having one of the disorders increases the odds of having the other approximately twofold. Covariates which were most likely to reduce the unadjusted association were those which were significantly associated with both MDD and CUD. However, no covariates significantly attenuated the OR between the two disorders.

A large section of the thesis is dedicated to twin model analyses, which decompose the covariance between CUD and MDD into genetic and environmental factors, and thereby inform the aetiology of this co-morbidity. However, twin models can only be fitted when it is established that the phenotypes of interest do significantly co-vary, which the current analyses have confirmed for this sample. Consequently, twin model analyses will be presented in the following chapters to examine the aetiology of the co-variance between the two phenotypes of interest.

5 Bivariate Twin Models of CUD and MDD

5.1 Introduction

Chapter 4 has established that the co-occurrence of Cannabis Use Disorder (CUD) and Major Depressive Disorder (MDD) was statistically significant in the current cohort of Australian twins, with an adjusted odds ratio of 1.92 (1.53–2.41; CUD as outcome) and 1.96 (1.57–2.45; MDD as outcome). This is in line with evidence from several twin, cross-sectional and longitudinal studies reviewed in Chapter 1. Despite consistent evidence of significant co-morbidity between MDD and CUD, there is limited understanding of the aetiological mechanisms underlying this relationship. Previous twin studies have not focused on the bivariate relationship between clinical levels of cannabis involvement and depression in general, and between MDD and CUD in particular. The twin model analyses presented in this chapter thoroughly examine the co-morbidity by focusing on genetic and environmental factors that might influence it.

As described in Chapter 3, twin models decompose the variance within a phenotype and the covariance between phenotypes into genetic and environmental factors: additive genetic (A), dominant genetic (D), shared environmental (C) and non-shared environmental (E) factors. As discussed in Chapter 1 (see Table 3), previous research has suggested that variance of both MDD and CUD are influenced by genetic and environmental factors. Pinpointing the sources of *covariance* is an important step toward understanding the aetiology of the co-morbidity, and targeting prevention or intervention efforts efficiently.

A bivariate correlated liabilities model was fitted to address two main aims: firstly, to ascertain which variance components influence CUD and MDD individually. The analyses examine whether the respective variances of both phenotypes are influenced by heritable factors in this cohort. This step is essential for establishing a link with previous literature and confirming that a behavioural genetic approach is appropriate to study the relationship between the phenotypes. The second aim is to examine how to what extent the covariance between the phenotypes can be

decomposed into genetic versus environmental factors. This will provide insight into the aetiological sources of their co-morbidity.

To fit the appropriate model, several steps had to precede the final model fitting. One prerequisite of bivariate twin model analysis is the presence of sufficient covariance between two phenotypes. Therefore, the first step was confirming that CUD and MDD were significantly correlated. Based on the results in Chapter 4, within-twin correlations between the phenotypes were expected to be significant. The second step was to correctly specify the model parameters: it needed to be decided which variance components would be estimated and how sex differences would be accounted for in the model. The likely importance of variance components was determined from between-twin correlations. Since previous literature has reported sex differences for the prevalence of both MDD and clinical levels of cannabis involvement (Kessler et al., 2005, 1994), sex differences were assessed using sex limitation models. These models examine whether any quantitative or qualitative sex differences need to be considered in the model. If quantitative sex differences are present, the twin model needs to specify that the same latent factors influence both sexes, but to different degrees. In the case of qualitative sex differences, the twin model should specify that different genetic and environmental factors influence males and females. Further details will be found in the methodology section.

In addition to bivariate twin analyses, discordant twin analyses were conducted. Any genetic or environmental correlation found in bivariate twin analyses provides information on the *source* of co-morbidity, i.e. genetic or environmental factors, but remains agnostic about the *process*. As outlined in Chapter 1, discordant twin analyses can disprove causality as a potential process underlying the co-morbidity between CUD and MDD. Therefore, a conditional logistic regression of MZ twins discordant for CUD and MDD was conducted. A significant association between CUD and MDD in MZ twins would suggest that causal influences between CUD and MDD are possible and can be investigated in further twin model analyses. These analyses will also be the first discordant twin analyses on CUD and MDD in males and females, using DSM-5 definitions (although excluding craving for CUD) from a general population twin sample.

In summary, this chapter will:

a) Test whether the covariance between CUD and MDD is significant, using *phenotypic correlations within individuals;*

b) Choose the appropriate twin model by examining:

    a. *Correlations between individuals* as indicators of variance components which need to be included in the model (A,C,D,E);

    b. *Sex limitation models* as indicators of quantitative and qualitative sex differences which need to be included in the model.

Once a significant covariance has been established and the appropriate twin model has been chosen, the main aims of this chapter are to:

    c. Determine which types of variance components significantly influence the variance and covariance of the phenotypes using *nested sub-models;*

    d. Estimate the heritability of CUD and MDD

    e. Estimate the genetic and environmental correlations between the phenotypes;

    f. Examine the possibility of causal influences between MDD and CUD through discordant twin analyses in MZ twins.

5.2 Methods

5.2.1 Sample

From a sample of 4131 twin pairs included in the Australian Twin Registry, 3824 twins and non-twin siblings born between 1972 and 1979 were interviewed on cannabis use, related drug use and other psychopathology. For full details of the sample see Chapter 2 and Lynskey et al. (2012).

After removing non-twin siblings and individuals of unknown zygosity, the analysis sample for the bivariate twin models comprised 3326 individuals: 976 monozygotic female (MZf), 490 monozygotic male (MZm), 741 dizygotic female (DZf), 373 dizygotic male (DZm) and 746 dizygotic opposite-sex (DZos) twins. The analyses sample for discordant twin analyses comprised 63 MZ pairs who were discordant for both CUD and MDD. The mean age of the sample was 32 years. Informed consent was obtained from all individual participants included in the study.

5.2.2 Measures

MDD and CUD were assessed using DSM-IV criteria, and recoded into DSM-5 variables as far as the available information allowed. Further details have been described in Chapter 2. In the analysis sample, 15.4% (11.9% of females, 22.6% of males) met the criteria for lifetime CUD and 26.1% (29.5% of females, 19.2% of males) met the criteria for lifetime MDD.

5.2.3 Statistical analyses

Data analysis for the bivariate correlated liabilities model was conducted using OpenMx (Neale et al., 2016) for R statistical software (R Core Team, 2014). Discordant twin analyses were performed in STATA (StataCorp, 2015).

5.2.3.1 The bivariate correlated liabilities model

This model has previous been introduced in Chapter 3 and will be described in further detail here. To estimate the heritability of both phenotypes and the genetic

and environmental correlations between them, a bivariate correlated liabilities model was fitted. The model contains separate latent genetic and environmental factors influencing each phenotype (see Figure 11). The heritability for CUD and MDD are estimated by the latent factor A, which is calculated by squaring the 'a' regression paths for each phenotype ('a^2'). The influence of other variance components is calculated by squaring their respective regression paths.

The genetic and environmental covariance components (between phenotypes) are calculated as a product of all factor-specific paths connecting the phenotypes. As an example, the covariance due to A is found by the following calculation: a x rA x a. Here, rA is the correlation between genetic factors influencing CUD and those influencing MDD.

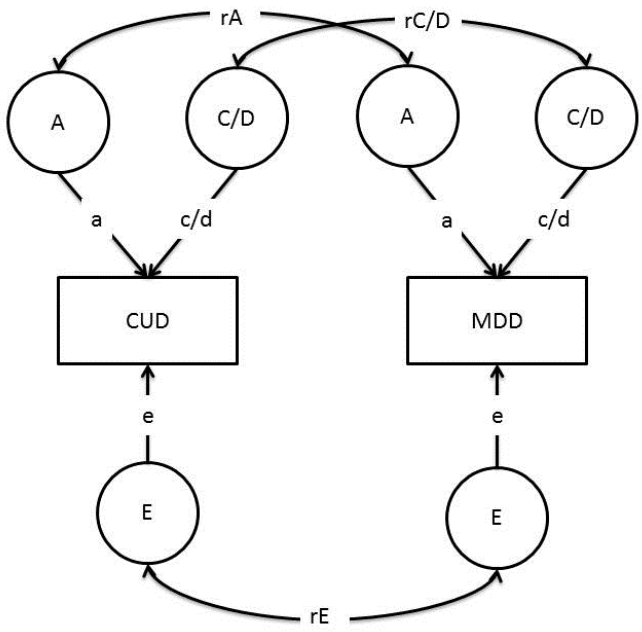

Figure 11. Example of bivariate correlated liabilities model for CUD and MDD. No quantitative or qualitative sex differences are taken into account.

Note. Upper case latter (A,C/D,E) = Latent additive genetic (A), shared environmental (C), dominance (D) and non-shared environmental (E) variance components
Lower case letter (a,c/d,e) = regression paths from latent variance component to observed phenotype (CUD or MDD)

5.2.3.2 Testing significance of covariance – tetrachoric correlations within individuals

Twin models can only be fitted when there is sufficient covariance between phenotypes, which can then be decomposed into genetic and environmental factors. It is also necessary to establish the magnitude of this covariance, to calculate the proportion which is explained by A, C/D and E. Therefore, the *phenotypic* tetrachoric correlation between MDD and CUD was computed. Since the OR between MDD and CUD was significant even after adjusting for confounding factors (see Chapter 4) , it was expected that the tetrachoric correlation between the phenotypes would be significant. This correlation is computed between phenotypes within individuals. It was not expected that this within-person correlation would vary by zygosity, but a variation by sex was possible and would be an indicator of sex differences.

5.2.3.3 Choosing the appropriate twin model

Given a significant phenotypic within-person correlation in males and females, several parameters need to be specified before fitting the final model. Primarily, it was important to decide whether a) the covariance between traits was likely to be influenced by heritable factors, b) sex differences needed to be accounted for and c) C (shared environment) or D (dominance) should be estimated in the model.

Prior to model fitting, the presence of sex differences and importance of variance components can be ascertained from *between-person* twin correlations. This informal initial assessment can be used to guide the selection of formal tests. Twin models informed by tetrachoric correlations are specified and then sub-models are tested formally to determine which sex differences and variance components should be included in the model most appropriate for the data.

5.2.3.4 Informal assessment of likely variance components and sex differences – Tetrachoric correlations between individuals

Firstly, between-individual correlations indicate how likely a trait is to be heritable. If MZ correlations are higher than DZ correlations, the variance in a trait or covariance between traits is likely to be influenced by genetic factors.

Secondly, these correlations provide evidence for the likely presence of sex differences. If male and female correlations differ or DZ opposite-sex twin correlations are lower than DZ same-sex correlations, this can indicate that sex differences are present in the sample and need to be examined. If there are differences in correlations between male and female same-sex twins, this suggests the presence of quantitative sex differences (Neale et al., 2006). If DZ opposite-sex correlations are lower than the correlations for DZ same-sex twins, males and females may be influenced by different factors and consequently display qualitative sex differences (Neale et al., 2006).

Thirdly, tetrachoric correlations between individuals indicate whether C or D should be estimated in the model. As explained in Chapter 3, the limited availability of parameters for the model-fitting process only allows three variance components to be estimated at any time. This leads to the choice between C and D in the model. This choice is made by comparing MZ to DZ correlations. Should MZ correlations be more than twice the size of DZ correlations, this can only be accounted for by the influence of dominance genetic factors. Therefore, C is often omitted from the model to estimate D.

Tetrachoric correlations within and between phenotypes for all zygosity groups were computed and all MZ and DZ correlations were compared to provide information on the issues above.

5.2.3.5 Formally examining sex differences

Formal tests of sex differences were examined based on the approach specified by Neale et al. (2006) and statistical details about the approach can be found in their

paper. Overall, it was examined whether two types of sex differences were present: quantitative and qualitative. Sex difference models were nested and compared to one another in terms of model fit. Further details were outlined in Chapter 3.

5.2.3.6 Formally testing the significance of variance components

The significance of variance components was tested formally by comparing model fit when each variance component was included in the model, to having a variance component fixed to zero. If model fit does not significantly deteriorate, the variance component fixed to zero is not thought to have a significant impact on the variance and/or covariance and was omitted from the model in the interest of parsimony.

5.2.3.7 Discordant twin analyses

As described in Chapter 2, discordant twin analyses were conducted using a conditional logistic regression. This regression is a form of logistic regression which allows for the matching of twins within a family. Consequently, the outcome of a conditional logistic regression is an OR between MDD and CUD within discordant twin pairs. The analyses only involved MZ twin pairs discordant for both MDD and CUD. Although MDD was analysed as the outcome and CUD as the predictor, any ordering of outcome or predictor would lead to the same result. Contrary to the bivariate twin models, sex and age did not need to be included as covariates, since they do not differ within twin pairs.

5.3 Results

5.3.1 The bivariate correlated liabilities model

5.3.1.1 Tetrachoric correlations within individuals – confirming significance of covariance

Tetrachoric correlations between MDD and CUD were examined *within* each individual, separately for males and females, in order to assess the magnitude of the relationship between the phenotypes, and to determine whether there are any sex

differences. These phenotypic correlations were *r* = .30 (95% CI = .21–.38) for females and *r* = .32 (95% CI = .21–.41) for males. These results indicated that there was sufficient covariance to decompose into genetic and environmental factors. The overlapping confidence intervals indicated that there were no sex differences in terms of the correlation between MDD and CUD within individuals.

5.3.1.2 Tetrachoric correlations between individuals – choosing the most appropriate model

In order to get an initial indication of the importance of different genetic and environmental factors, the importance of sex differences, and hence the appropriate twin model, within- and cross-trait tetrachoric correlations *between* twins of different zygosities were computed.

Table 20. Tetrachoric correlations between CUD and MDD by zygosity.

Zygosity	CUD - CUD	MDD - MDD	CUD - MDD
MZf	.84 (.71–.91)	.46 (.31–.60)	.24 (.12–.37)
DZf	.50 (.26–.69)	.08 (.04–.27)	.15 (-.02–.31)
MZm	.70 (.48–.84)	.43 (.12–.67)	.15 (-.04–.33)
DZm	.42 (.11–.67)	.24 (-.15–.57)	.15 (-.10–.39)
DZos	.17 (-.07–.39)	.22 (-.00–.43)	.17 (.01–.20)

Note. CUD-CUD: correlation between twin 1 and twin 2 CUD
MDD-MDD: correlation between twin 1 and twin 2 MDD
CUD-MDD: cross-twin cross-trait correlation

Table 20 shows the within-trait and cross-trait correlations between twins for MDD and CUD, by zygosity. The results in this table indicate:

1. For CUD, the differences in within-trait correlations between MZ and DZ twins suggest that A may explain a substantial portion of the variance in males (around 58%) and females (around 68%). An ACE, rather than ADE model is likely to be most adequate, since neither MZf nor MZm correlations were over twice as large as the respective DZ correlations. The correlation between opposite sex DZ twins was

somewhat, but not substantially, lower than the DZ same-sex correlation. This may be an indicator of qualitative sex differences in CUD.

2. For MDD, A also seems to play an important role for males and females, possibly explaining around 38% and 76% of the variance, respectively. As the MZf within-trait correlation is substantially larger than two times the DZf correlation, an ADE model may be the more appropriate model for MDD. The correlations of DZ opposite sex twins were not significantly lower than those of same-sex twins, which indicated that qualitative sex differences may not play an important role with respect to MDD.

3. The cross-twin cross-trait correlations for MDD and CUD indicate that additive genetic factors may influence the relationship between MDD and CUD in females, but not in males. The MZm and DZm correlations are equal. In females, A is not likely to play a substantial role, because confidence intervals between MZf and DZf twins are overlapping and the estimated portion of covariance explained by A would be 18%.

Overall, an ACE model was fitted. Although the difference between MZf and DZf twin correlations for MDD was larger than two, all other correlations were suggestive of an ACE model. Furthermore, the results were indicative of possible sex differences, which were tested in the next step.

5.3.1.3 Sex differences – formal tests

Several sex limitation models were fitted, all of which were bivariate correlated liabilities models.

Five different models were used to comprehensively examine quantitative and qualitative sex differences. Model 1 assumed quantitative sex differences for all variance components and qualitative sex differences for A. Model 2 estimated quantitative sex differences for all variance components and qualitative sex differences for C. Under model 3, only quantitative sex differences were estimated. Model 4 assumed no quantitative sex differences, i.e. all regression paths from the variance components to the phenotypes, and correlation paths between variance

components were equated across sexes. Model 5 tested whether thresholds could be equated for males and females. Model fit comparisons are summarised in Table 21.

Table 21. Model fit comparison between sex limitation models for CUD and MDD.

Base	Comparison	EP	Likelihood	DF	AIC	*P*
Model 1		30	6405.63	6564	-6722.37	-
Model 1	Model 3	26	6405.96	6568	-6730.05	.988
Model 2		30	6405.67	6564	-6722.33	-
Model 2	Model 3	26	6405.96	6568	-6722.33	.991
Model 3		26	6405.96	6568	-6730.05	-
Model 3	Model 4	17	6414.30	6577	-6739.70	.499
Model 4		17	6414.30	6577	-6739.70	-
Model 4	Model 5	13	6511.85	6581	-6650.15	<.001

The best fitting and most parsimonious model was model 4, which assumed differences in thresholds (i.e. prevalence), but no quantitative or qualitative sex differences.

5.3.1.4 Variance components – formal tests

Since estimates for shared environmental influences were low in model 4, an AE model was fitted to the data. Removing C from the model did not significantly deteriorate model fit (see Table 22).

Table 22. Model comparison results between bivariate correlated liability models, with and without shared environmental factors (C).

Base	Comparison	EP	Likelihood	DF	AIC	*p*
ACE		17	6415.08	6577	-6738.92	
ACE	AE	14	6415.08	6580	-6744.92	1

In fact, the estimates of C were so small in the ACE model that the likelihoods of the ACE and AE model are identical. The resulting model is summarised in Figure 12. In the AE model the parameter estimates remained the same as in the ACE model outlined above.

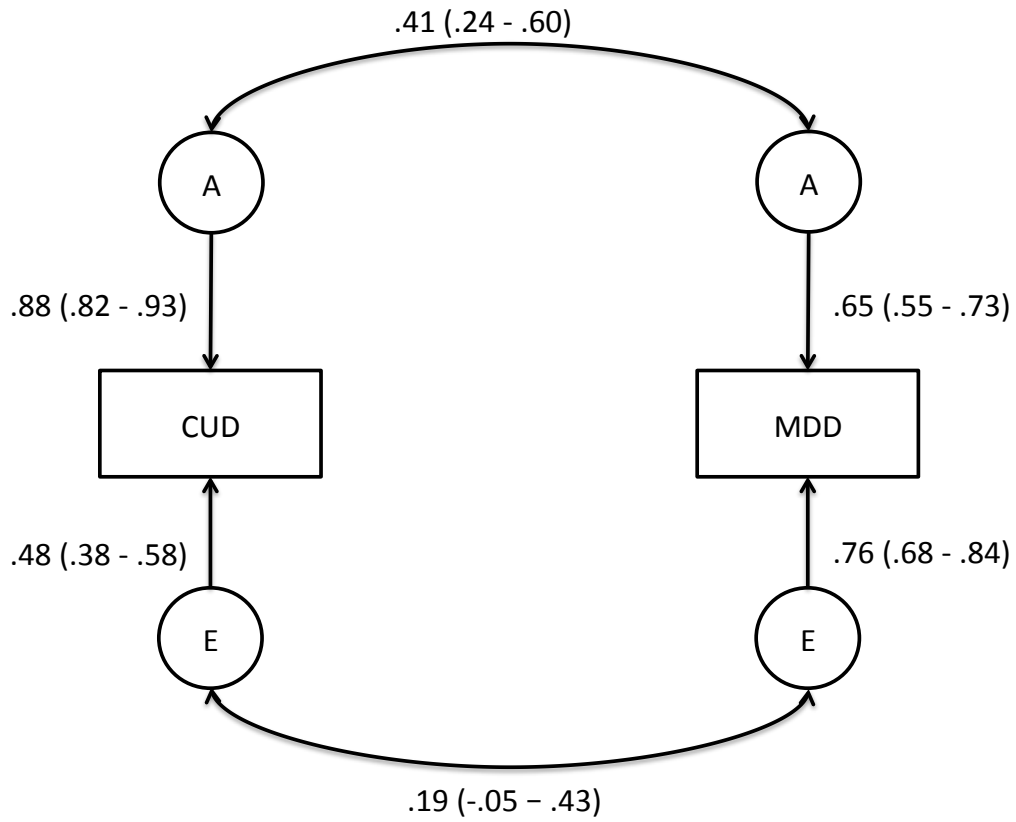

Figure 12. Bivariate AE correlated liabilities model for CUD and MDD with regression coefficients and confidence intervals

In the final model, genetic factors explained 77% (95% CI = 67%–85%) of the total variance in CUD, and non-shared environmental factors the remaining 23% (95% CI = 16%–27%). For MDD, genetic factors explained 42% (95% CI = 30%–54%) and non-shared environmental factors 58% (95% CI = 46%–69%) of the total variance. The genetic correlation between the two phenotypes was r_g = .41 (95% CI = .24–.60) and the non-shared environmental correlation r_e = .19 (95% CI = -.05–.42). Genetic factors explained around 77% of the total phenotypic correlation.

5.3.2 Discordant twin analyses

A conditional logistic regression of MZ twin pairs discordant for CUD showed that MZ twins with CUD had significantly elevated rates of MDD (46.0%) relative to their co-twins who did not have CUD (28.12%; OR = 2.83, 95% CI = 1.12–7.19; N = 63 MZ pairs). CUD explained 17% of the variance in MDD (*pseudo-*R^2 = .17).

5.4 Discussion

5.4.1 The bivariate correlated liabilities model

The first aim of this chapter was to establish whether A, C/D or E influence the variance of CUD and MDD. In the current cohort, heritability was estimated at 0.77 for CUD and 0.42 for MDD. Aside from additive genetic factors (A), non-shared environmental (E), but not shared environmental (C) factors influenced the phenotype-specific variance. These results are in line with previous research for MDD (Kendler, Gatz, et al., 2006; Sullivan et al., 2000), and within the range of estimates obtained for Cannabis Abuse and Dependence (Kendler, Aggen, Tambs, & Reichborn-Kjennerud, 2006; Verweij et al., 2010).

Establishing the genetic and environmental correlations between the phenotypes was the second aim, and these correlations were estimated to be 0.41 for the genetic factors and 0.19 for the non-shared environmental factors. In the most parsimonious model, shared environmental factors were not included. While the genetic correlation was statistically significant, the shared environmental correlation was not. The genetic correlation also explained most of the phenotypic correlation (77%), which suggests that the co-morbidity between CUD and MDD may be driven mostly by shared genetic factors.

Additionally, in the current study no significant quantitative or qualitative sex differences emerged. The only sex differences found in the current analyses were with respect to prevalence of both MDD and CUD, which is consistent with previous literature (Kessler et al., 2005, 1994)

Since this is the first study to examine a bivariate correlated liabilities model in CUD (with the exception of craving) and MDD, and report correlations for all genetic and environmental factors, there is limited twin literature with which these results can be compared. Fu et al. (2002) do not report the magnitude of the genetic correlation between MDD and cannabis dependence in their sample. For lifetime cannabis dependence and MDD, Lynskey et al. (2004) report a genetic correlation of 0.44 (95% CI = 0.17–1.00) for males and 0.69 (95% CI = 0.30–1.00) for females. The

current results are within the confidence interval for both. However, the current study looked at lifetime CUD, not cannabis dependence and is therefore not entirely comparable. To the best of the author's knowledge, there are currently no other twin studies which have provided comparable twin estimates.

It would be a valuable to compare the magnitude of the genetic correlation estimates from the bivariate twin study in this thesis with those estimated using molecular genetic data. Molecular genetic studies also suggest that there is significant genetic overlap between MDD and CUD (Carey et al., 2016; Demontis et al., 2018; Sherva et al., 2016), although they do not report on the magnitude of this association, only that it is significantly non-zero. While Hodgson et al. (2016) do report a genetic correlation (ρ_g = 0.42, *p* = .036, SE = 0.20), they looked at the relationship between lifetime cannabis use and MDD, not CUD and MDD. Given that large-scale genomewide association studies (GWAS) have been conducted for CUD and MDD already, a simple approach for estimating the magnitude of SNP-based genetic correlation would be to use a GWAS summary statistic based method such as LD-Score Regression (Bulik-Sullivan, Finucane, et al., 2015; Bulik-Sullivan, Loh, et al., 2015). Alternatively, with individual-level genetic and phenotypical data, the genetic correlation could be estimated using the GREML (genomic-relatedness-based restricted maximum-likelihood) method implemented in the software GCTA (genome-wide complex trait analysis; Yang et al., 2010). Nonetheless, these previous genotype-based studies in combination with the findings in this thesis provide substantial evidence for a significant genetic correlation between CUD and MDD.

Overall, the bivariate twin analyses suggest that CUD and MDD share overlapping genetic factors, which is a plausible explanation for their-co-morbidity: although the specific site is unknown, studies on humans and animals point toward the endocannabinoid system as a possible site of overlapping genetic vulnerability. As mentioned in Chapter 1, this is not only the primary site of the neurochemical effects of cannabis but is also thought to be involved in mood regulation (Ashton & Moore, 2011). Several studies have linked the CB1 receptor, and its encoding gene CNR1, to both THC and mood regulation, (see Agrawal and Lynskey, 2014 for a review), making it a likely candidate for shared genetic vulnerabilities between CUD and

MDD. Additionally, overlapping genetic factors may be due to a genetically influenced covariates.

5.4.1.1 Strengths and limitations

The primary advantage of fitting a correlated liabilities model is its power to investigate the *source* of covariance between two phenotypes without presupposing a process underlying co-morbidity or any direction of effect. Different models, such as Cholesky decompositions used in one study on MDD and cannabis dependence (Fu et al., 2002) make assumptions about both sources and processes, because they require specifying a variable order (i.e. CUD before MDD or MDD before CUD). The estimates and interpretations of the model change depending on this variable order. In contrast, correlated liabilities model allow remaining agnostic about variable order and therefore the direction of any causal or non-causal effects.

However, this means that correlated liabilities models cannot inform about the likely process of co-morbidity. Discordant twin models and analyses in Chapter 6, have been carried out to overcome this limitation.

5.4.2 Discordant twin analyses

Discordant twin analyses showed that causal processes could not be excluded as an explanation for the correlation between MDD and CUD, because MZ twins with CUD were significantly more likely to display symptoms of MDD than their co-twins without CUD. Therefore, there is sufficient reason to continue exploring causality in further analyses. The results are in contrast to study results in Lynskey et al. (2004) and in line with those in Lin et al. (1996), but overall they are not entirely comparable to either due to differences in variable definitions and samples.

Discordant twin results are difficult to interpret due to the small number of MZ twin pairs discordant for both CUD and MDD. The sample size is characteristic of analyses on single cohorts, because – even in large cohorts – twin pairs discordant for both phenotypes are rare. For example, Lynskey et al.'s (2004) analysis sample comprised 87 twin pairs when they examined the association between MDD before

age 17 on later onset cannabis dependence. The vast reduction from total to analysis sample occurs because twins need to be discordant for both phenotypes.

To conclusively examine whether causal processes can be excluded using the discordant twin method, larger sample sizes would be particularly critical with this type of analysis. A recent study by Agrawal et al. (2017) has been conducted on a large sample of twins, but the results cannot be compared as the current cohort was included in their analyses and they investigated frequent cannabis use (more than 100 times used) rather than clinical definitions. For the purposes of the current study, the main intention was establishing that causal processes could not be ruled out within the current data set.

5.4.3 Conclusion and further analyses

In conclusion, overlapping genetic risk factors between CUD and MDD are both a plausible explanation of co-morbidity and have been found in the current study, as well as others. The correlated liabilities model fitted in the presented analyses is a model which is *agnostic* toward the specific process by which genetic factors contribute to the co-morbidity. Overlapping risk factors may mean that these risk factors simultaneously give rise to both CUD and MDD, but there are various other possible mechanisms. Discordant twin analyses suggested that causal mechanisms may be among them. Thirteen of those aetiological mechanisms, including but not limited to causality, will be explored in Chapter 6.

6 Co-morbidity models of CUD and MDD

6.1 Introduction

The previous results chapters have demonstrated that Major Depressive Disorder (MDD) and Cannabis Use Disorder (CUD) are significantly co-morbid in this sample (OR = 2.23, 95% CI = 1.84–2.70), may be causally related and their covariance is primarily explained by overlapping genetic factors (*r* = .41, 95% CI = .24–.60). However, the aetiological processes by which these factors influence the co-morbidity remain unclear. This chapter aims to comprehensively investigate competing models of co-morbidity and will fit Neale and Kendler's (NK; 1995) 13 co-morbidity models, which were based on the work of Klein and Riso (1993). Each model and each class of models makes different assumptions about the aetiological mechanisms that lead to co-morbidity. Four broad classes are examined: single liability, independent liability, multiformity and correlated liabilities. No other twin model approach examines such a large variety of model classes.

If the co-morbid form arises from a *single liability* shared by CUD and MDD, the diagnostic boundary between MDD and CUD may have been artificially drawn, and they could be alternate forms of the same disorder. This model is not likely to explain the co-morbidity between MDD and CUD well, because it assumes a full overlap between the disorders. If this was the case, one would expect a much higher rate of co-morbidity found in this and other samples, and a significantly stronger genetic correlation between the disorders in the bivariate twin model (Chapter 5).

Alternatively, the liability of the co-morbid form may be entirely *independent,* and 'co-morbid CUD and MDD' could be a third disorder, unrelated to the pure forms of MDD and CUD. This model may not be a likely explanation for the observed co-morbidity either, since it assumes that the liabilities of MDD and CUD, in their pure form, are entirely unrelated. Given the likely overlap in neurobiology responsible for mood regulation and the psychoactive effects of cannabis (see Chapter 1) , this model is unlikely to be supported by the data.

Multiformity models suggest that the risk factors for CUD and MDD are unrelated, but once certain thresholds on the liability of one disorder are crossed, the risk of symptoms of the other disorder increases sharply. In other words, MDD and CUD influence each other in a discontinuous way, only once certain levels of risk are reached. This sudden increase in co-morbidity at high levels of risk is compatible with findings from longitudinal studies, which produce mixed results when examining non-clinical definitions of cannabis involvement and depression, but report the highest associations between clinical levels of both (Bovasso, 2001; Lev-Ran et al., 2014; Marmorstein & Iacono, 2011; Pacek et al., 2013). Multiformity models are therefore plausible for CUD and MDD.

Correlated liabilities models assume that liabilities between two disorders are related continuously, and aetiological factors overlap. Any change in risk for one disorder is accompanied by a change in risk for the other disorder, whether this is due to shared risk factors or causality. Correlated liabilities models without the assumption of causality are likely to fit the data well, since they are similar to those fit in Chapter 5. Since discordant twin analyses suggest that causal influences between CUD and MDD are possible, causal models may also fit the observed data well.

In addition, the models test whether the co-morbidity observed in the population has occurred by chance.

The NK co-morbidity models have been used to examine the co-morbidity between a range of other substance-use and psychiatric phenotypes (Agrawal et al., 2007, 2004, 2010). To mention one applied example of these twin models, Agrawal et al. (2007) investigated the 'gateway hypothesis', according to which cannabis use precedes and increases the risk of other illicit drug use (Kandel, Yamaguchi, & Chen, 1992). After fitting all 13 models to a sample of 4152 Australian twins, the correlated liabilities model provided the best fit to the data, with both genetic and environmental factors being highly correlated between cannabis and illicit drug use. This result suggests that cannabis and other illicit drug use co-occur due to shared risk factors influencing both. However, both the reciprocal causation and unidirectional causation (cannabis use causes other illicit drug use) models fit the data well. Consequently, some causal influence could not be excluded, and there was evidence for the gateway hypothesis (i.e. cannabis use causing other illicit drug use). This conclusion is supported by the

finding that high-risk female users of cannabis were at an increased likelihood to use other illicit drugs, irrespective of their individual liability to use other illicit drugs (*extreme multiformity* of cannabis use). In summary, this study has highlighted sex differences and strongly narrowed down the range of possible forms of co-morbidity between cannabis and other illicit drug involvement. Further genetic models can be applied to differentiate between the highlighted models.

Previous literature suggests that twin models of co-morbidity would also be a useful tool to study the relationship between cannabis involvement and depression (Agrawal & Lynskey, 2014), since both MDD (e.g. Sullivan et al. 2000; Kendler et al. 2006) and cannabis dependence (Lynskey et al., 2002; Verweij et al., 2010) are influenced by genetic factors. Chapter 5 has demonstrated that this holds true for MDD and CUD in this sample as well.

To date, no study has examined all 13 models with respect to MDD and CUD, although previous longitudinal and twin studies have produced conflicting findings regarding the relationship between these phenotypes. Therefore, analyses in the current chapter aimed to fit all 13 NK co-morbidity models to examine the relationship between CUD and MDD.

6.2 Methods

6.2.1 Participants

The sample has been described in detail in Chapter 2. The analyses presented in this chapter were conducted on twins only and required complete data from each twin pair for both phenotypes. The sample size is therefore somewhat different from that of previous chapters: 2410 individual twins were included in the analysis sample, 565 (396 female, 169 male) complete MZ pairs and 640 (298 female-female, 118 male-male and 224 female-male) complete DZ twin pairs.

6.2.2 Measures

MDD and CUD were assessed using DSM-IV criteria, and recoded into DSM-5 variables as far as the available information allowed: the definition of CUD included all

DSM-5 criteria except for craving. Further details have been described in Chapter 2. In the analysis sample, 14.75% individuals met the criteria for lifetime CUD and 25.55% met the criteria for lifetime MDD.

6.2.3 Statistical models

A summary of all models can be found in Table 23. Each model makes different assumptions about the way in which co-morbid cases arise. A detailed statistical discussion of the models can be found in Chapter 3, as well as in Neale and Kendler (1995) and Rhee et al. (2004). This chapter aims to provide an explanation specifically applied to MDD and CUD.

Table 23. Summary and interpretation of Neale and Kendler (1995) models of co-morbidity.

Nr.	Model	Sub-models	Description applied to CUD and MDD
1	Alternate Forms		Single liability: after crossing a common threshold, some develop CUD, some MDD. MDD and CUD are alternate forms of the same disorder.
2	Three independent disorders		The pure forms are unrelated disorders. There are three independent liabilities for CUD, MDD and co-morbid CUD with MDD.
	Multiformity		The liabilities for CUD and MDD are unrelated. CUD discontinuously increases the risk of MDD symptoms, and vice versa when thresholds are crossed. Random multiformity assumes one, extreme multiformity two thresholds.
3	Random Multiformity (RM)		Assumes a single threshold within one disorder (e.g. CUD), above which the risk to develop symptoms of the other disorder (e.g. MDD) suddenly increases. This model allows for *both* disorders to increase the risk of symptoms of the other.
4		RM of MDD	Being above the threshold for MDD risk leads to a sudden increase in risk for symptoms of CUD, even when below the threshold for CUD.
5		RM of CUD	Being above the threshold for CUD risk leads to a sudden increase in risk for symptoms of MDD, even when below the threshold for MDD.
6	Extreme Multiformity (EM)		There are two distinct thresholds for both disorders. Crossing the first threshold leads to the pure form of a disorder. The second threshold allows for individuals with high amounts of risk factors. Individuals will be at increased risk for symptoms if they are above the second threshold (at increased risk) for either disorder.
7		EM of MDD	Being above the first threshold for MDD risk only leads to MDD. A proportion of high-risk individuals with MDD (above the second threshold) develop CUD symptoms, even when below the first threshold for CUD.
8		EM of CUD	Being above the first threshold for CUD risk only leads to CUD. A proportion of high-risk individuals with CUD (above the second threshold) have MDD symptoms, even when below the first threshold for MDD risk.
9	Correlated liabilities		Correlation between latent genetic and environmental influences on CUD and MDD gives rise to co-morbidity.
10		Reciprocal Causation	The liability for CUD has causal influence on the liability for MDD, and vice versa.
11		Unidirectional: MDD to CUD	The liability for MDD has a causal influence on the liability for CUD.
12		Unidirectional: CUD to MDD	The liability for CUD has causal influence on the liability for MDD.
13		Chance	Co-morbid CUD and MDD occur due to chance alone.

Because both phenotypes were coded as binary variables, the foundation of each co-morbidity model was a normal liability threshold model, similar to that in Chapter 5.

Similar to widely used liability threshold models, all models estimate genetic (A), shared (C) and non-shared environmental (E) factors. D was not estimated in the current sample, because the difference between MZ and DZ correlations indicated an ACE model. The rationale for this decision was based on between-person correlations, and explained in Chapter 3. Despite these similarities, there are several important differences between models:

1. The models differ in the number of liability distributions they assume. For instance, the alternate forms model assumes that both phenotypes arise from *one* distribution of liability. In contrast, the three independent disorders model assumes that there are *three* underlying liability distributions. Two of those give rise to the pure forms of the phenotypes, and one gives rise to the co-morbid form.

2. The models differ in the way in which the above-mentioned liabilities produce the phenotype. For example, in the alternate forms model, an individual develops co-morbid CUD and MDD by crossing the threshold on the shared liability distribution. However, in the three independent disorders model, an individual can develop CUD and MDD if they cross the threshold on the CUD-specific and MDD-specific distribution at the same time, or if they do so on the liability distribution for the co-morbid form.

3. The extreme multiformity model also differs from all others in the number of thresholds it assumes. Under the assumptions of this model, each liability distribution has two thresholds. If an individual crosses the first threshold, they only develop the pure form of the disorder. Crossing the second threshold means that the individual develops the co-morbid form. Consequently, co-morbidity arises if an individual crosses the first threshold on both liability distributions, the second threshold on one liability distribution (e.g. CUD), and/or the other distribution (e.g. MDD).

6.2.4 Data analysis

Data analysis was conducted using OpenMx (Neale et al., 2016) for R statistical software (R Core Team, 2014). The input to each model was a frequency table, which summarised the number of twin pairs fitting into 10 MDD-CUD co-morbidity categories (see Table 24). Each twin pair member was assigned to one of four disease state categories: MDD but no CUD (i.e. 1 0), no MDD, but CUD (i.e. 0 1), both MDD and CUD (i.e. 1 1), and neither (i.e. 0 0). Thereafter, the twin disease states were combined (i.e. 0 0 0 1). Although there are 16 different combinations of co-twin disease states, information about twin order was disregarded to avoid low cell counts. For instance, '0 0 1 0' (see Table 24) is a category that contains cases where twin 1 only (i.e. 1 0 0 0) *or* twin 2 only (i.e. 0 0 1 0) was affected by MDD. Subsuming all replicating disease states resulted in 10 categories.

Table 24. Numbers of twin pairs in co-morbidity status categories. Used as input for all co-morbidity models.

	Twin 1		Twin 2		MZ	DZ
	MDD	CUD	MDD	CUD		
1.	0[a]	0	0	0	298	277
2.	0	0	0	1[b]	28	73
3.	0	0	1	0	114	145
4.	0	0	1	1	17	35
5.	0	1	0	1	16	10
6.	0	1	1	0	6	21
7.	0	1	1	1	16	23
8.	1	0	1	0	47	33
9.	1	0	1	1	12	18
10.	1	1	1	1	11	5
Total					565	640

[a] 0 = unaffected
[b] 1 = affected

For every model, the number of twin pairs expected in each of the 10 categories was based on the assumptions of the model. Maximum likelihood estimation was used to minimise the difference between the observed number of cases in each co-morbidity category and the expected number according to the model. A chi-squared goodness-of-fit (**χ2**) test compared these observed and expected values and indicated model fit. The *p*-value of the **χ2** test was used to reject models whose predicted data was significantly different from the observed data. The best fitting and most parsimonious

model was chosen based on the Akaike Information Criterion (AIC; Akaike 1987). According to Burnham and Anderson (2002), an AIC difference of three or greater indicates that the model with the lower AIC has substantially more support.

6.3 Results

The model-fitting results are summarised in Table 25. In addition to the 13 co-morbidity models, a saturated model (model containing an equal number of parameters and data points) based on twin correlations was included for comparison. Five models can be rejected due to the large, statistically significant differences between the observed cell counts within the co-morbidity categories (see Table 24) and the cell counts expected under the chance, alternate forms, three independent disorders, RM of MDD, and EM of MDD models.

Table 25. Co-morbidity model fit statistics, indicating fit of each model and differences between cell counts predicted under the model and observed call counts

Model	χ2	Df	p	AIC
Saturated Model	13.16	11	.283	-8.84
1. Alternate Forms	96.31	14	<.001	68.31
2. Three Indep. Disorders	32.90	8	<.001	16.90
3. Random Multiformity	15.10	10	.129	-4.90
4. RM of MDD	27.04	11	.004	5.04
5. RM of CUD	15.46	11	.162	-6.54
6. Extreme Multiformity	16.32	10	.091	-3.68
7. EM of MDD	28.23	11	.003	6.23
8. EM of CUD	19.50	11	.053	-2.50
9. Correlated Liabilities	15.21	9	.085	-2.79
10. Reciprocal Causation	15.23	10	.124	-4.77
11. MDD causes CUD	17.86	11	.085	-4.14
12. CUD causes MDD	15.50	11	.161	-6.50
13. Chance	59.46	12	<.001	35.46

The only models that do not have substantially less support than the saturated model (i.e. an AIC difference larger than 3) are the RM of CUD model (model 5) and the CUD causes MDD model (model 12; see Figure 13a and b), with AIC differences of 2.30 and 2.34 respectively. Both models have substantially more support than the correlated liabilities model.

1a. *Random Multiformity of CUD* **1b.** *Causation – CUD causes MDD*

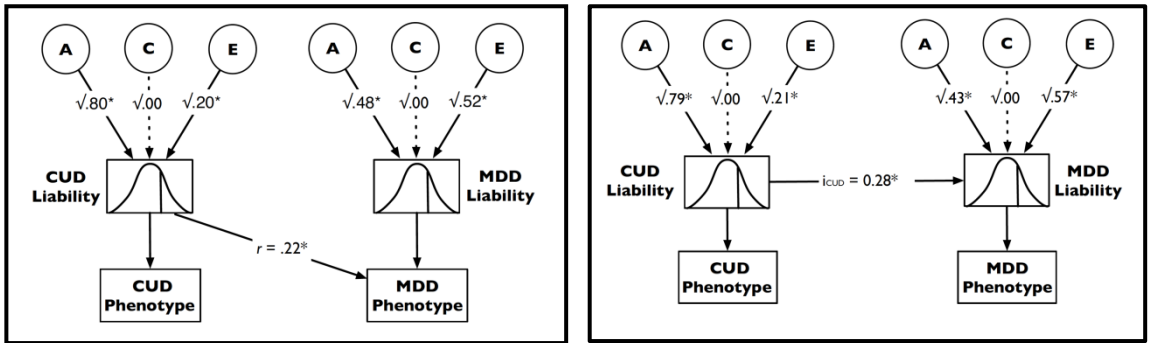

Figure 13a and b. Parameter estimates from best fitting co-morbidity models: Random Multiformity of CUD and CUD causes MDD.

Note. *r* = probability of MDD phenotype if above threshold on CUD liability
i_{CUD} = regression coefficient
* = significant at the 0.05 level

These two best fitting models are, however, not substantially different from some of the models within their class. The RM of CUD model is not substantially different from the Random Multiformity model. The CUD causes MDD model is not substantially different from the MDD causes CUD and Reciprocal Causation model. Additionally, both models do not substantially differ from the Extreme Multiformity model.

In the best fitting models both CUD and MDD are influenced by genetic and non-shared environmental factors. In the case of CUD, 79–80% of the total variance is estimated to be explained by genetic factors and 20–21% by non-shared environmental factors. For MDD, 43–48% of the total variance is explained by genetic factors, and 52–57% by non-shared environmental factors. Model fit did not significantly deteriorate when C was dropped from both models. The parameter estimates from all models can be found in Table 26.

Table 26. Parameter estimates of all co-morbidity models for MDD and CUD.

	Model						
	1	2	3	4	5	6	7
a^2_{MDD}	-	.46 (.23 - .56)	.48 (.20 - .60)	.46 (.23 - .57)	.48 (.19 - .60)	.48 (.17 - .59)	.46 (.23 - .56)
c^2_{MDD}	-	.00 (.00 - .17)	.00 (.00 - .22)	.00 (.00 - .17)	.00 (.00 - .23)	.00 (.00 – .35)	.00 (.00 - .45)
e^2_{MDD}	-	.54 (.44 - .66)	.52 (.40 - .65)	.54 (.43 - .66)	.52 (.40 - .64)	.52 (.41 - .65)	.54 (.44 - .66)
a^2_{CUD}	-	.76 (.45 - .87)	.82 (.52 - .90)	.86 (.53 - .94)	.80 (.50 - .87)	.74 (.41 - .83)	.76 (.45 - .87)
c^2_{CUD}	-	.03 (.00 - .29)	.00 (.00 - .24)	.03 (.00 - .31)	.00 (.00 - .24)	.01 (.00 - .28)	.03 (.00 - .29)
e^2_{CUD}	-	.21 (.13 - .31)	.18 (.10 - .29)	.12 (.06 - .21)	.20 (.13 - .31)	.25 (.17 - .36)	.21 (.13 - .31)
a^2_{shared}	.66 (.66 - .66)	.53 (.00 - 1.00)	-	-	-	-	-
c^2_{shared}	.34 (.34 - .34)	.02 (.00 - 1.00)	-	-	-	-	-
e^2_{shared}	.00 (.00 - .00)	.45 (.00 - 1.00)	-	-	-	-	-
p^a	.44 (.44 - .44)	-	.02 (.02 - .08)	.10 (.06 - .14)	-	-	-
r^b	.27 (.27 - .27)	-	.20 (.09 - .27)	-	.22 (.15 - .29)	-	-
$i_{MDD}{}^c$	-	-	-	-	-	-	-
$i_{CUD}{}^d$	-	-	-	-	-	-	-
$t_{MDD}{}^e$	-	0.63	0.71	0.64	0.72	0.69; 2.21	0.64; 1.92
$t_{CUD}{}^f$	-	0.96	1.01	1.06	0.99	1.04; 1.98	1.07
t_{shared}	-0.31	1.13	-	-	-	-	-

Table continues on next page

	Model					
	8	9	10	11	12	13
a^2_{MDD}	.48 (.17 - .59)	.45 (.18 - .56)	.44 (.19 - .55)	.47 (.22 - .57)	.43 (.18 - .54)	.46 (.23 - .56)
c^2_{MDD}	.00 (.00 - .08)	.01 (.00 - .21)	.00 (.00 - .14)	.00 (.00 - .19)	.00 (.00 - .18)	.00 (.00 - .18)
e^2_{MDD}	.52 (.41 - .65)	.56 (.44 - .67)	.56 (.45 - .70)	.53 (.43 - .65)	.57 (.46 - .70)	.54 (.44 - .66)
a^2_{CUD}	.74 (.41 - .83)	.79 (.48 - .87)	.80 (.48 - .88)	.81 (.52 - .88)	.79 (.79 - .87)	.76 (.45 - .87)
c^2_{CUD}	.01 (.00 - .28)	.01 (.00 - .26)	.00 (.00 - .03)	.00 (.00 - .23)	.00 (.00 - .26)	.03 (.00 - .29)
e^2_{CUD}	.25 (.17 - .36)	.20 (.13 - .31)	.20 (.12 - .30)	.19 (.12 - .30)	.21 (.13 - .32)	.21 (.13 - .31)
a^2_{shared}	-	.19 (-.03 - .31)	-	-	-	-
c^2_{shared}	-	.01 (-.07 - .17)	-	-	-	-
e^2_{shared}	-	.08 (-.01 - .17)	-	-	-	-
p^a	-	-	-	-	-	-
r^b	-	-	-	-	-	-
$i_{MDD}{}^c$	-	-	.07 (-.19 - .32)	.26 (.18 - .35)	-	-
$i_{CUD}{}^d$	-	-	.21(-.05 - .52)		.28 (.20 - .38)	
$t_{MDD}{}^e$	0.72	0.64	0.66	0.64	0.66	0.63
$t_{CUD}{}^f$	1.00; 1.81	0.99	1.01	1.03	.0.99	0.96
t_{shared}	-	-	-	-	-	-

Note. Model numbers refer to the co-morbidity models outlined in Table 23. Shared a^2, c^2 and e^2 refer to the single shared liability in model 1, third independent liability in model 2, or Cholesky paths from MDD to CUD in model 9.

[a] *p* = probability of CUD if above threshold on the MDD liability (models 3 -5), or above threshold on the shared liability (model 1)

[b] *r* = probability of MDD if above threshold on the CUD liability (models 3 -5), or above threshold on the shared liability (model 1)

[c] i_{CUD} **=** regression coefficient from CUD to MDD,

[d] i_{MDD} = regression coefficient from MDD to CUD

[e] t_{MDD} = threshold for MDD

[f] t_{CUD} = threshold for CUD

6.4 Discussion

To the author's knowledge, this study is the first to fit the 13 co-morbidity models proposed by Neale and Kendler (1995) to Cannabis Use Disorder and Major Depressive Disorder. The two best-fitting models were the Random Multiformity of CUD and CUD causes MDD models. Both models fit substantially better than the Correlated Liabilities model, and not substantially worse than the Saturated Model. In addition, five models could be statistically rejected: the Alternate Forms, Chance, Three Independent Disorders, RM of MDD and EM of MDD models. The heritability estimates in the best fitting models range from 79%–80% for CUD and 43–49% for MDD.

6.4.1 Model-fitting

These model-fitting results suggest that the direction of effect goes from CUD to MDD. Firstly, both the RM of MDD and EM of MDD models can be statistically rejected. It seems plausible, therefore, that the fit of the bi-directional random multiformity and extreme multiformity models is driven by the paths they have in common with the RM of CUD and EM of CUD models, respectively. Secondly, the CUD causes MDD model fits better than the MDD causes CUD model. Although this difference is not substantial, the fit of the MDD causes CUD model may reflect that direction of causation models are difficult to distinguish when modes of inheritance of the disorders are similar (Heath et al., 1993). In the current study, this may be because both disorders are mainly influenced by A and E, rather than by different aetiological factors (e.g. A C E vs. A E). Lastly, the MDD causes CUD model, along with all other models with a direction of effect from MDD to CUD, was a substantially poorer fit than the saturated model.

It is unclear, however, which of the two best-fitting models is more likely. The CUD causes MDD model assumes that the liability to develop MDD symptoms increases *continuously* as the risk of CUD increases. The threshold in this model does not equal a sudden increase in risk, which means that even sub-threshold increases in liability to CUD have a causal influence on the liability to develop MDD (Rhee et al.,

2004). On the other hand, the RM of CUD model assumes that the risk of MDD symptoms increases *discontinuously*, once the threshold on the CUD liability has been passed (i.e. an individual has reached a liability high enough to develop the disorder). An additional difference between the models is their assumption about aetiological processes. The causal model assumes that any causal processes occur at the level of the liability (Rhee et al., 2004), while the RM models remain agnostic about the way in which one disorder leads to symptoms of the other (i.e. via risk factors or at the phenotypic level).

Despite some differences, the RM of CUD and CUD causes MDD models are not incompatible. Causality may play a role, and the good fit of the RM of CUD model may indicate that the causal influences on the risk of MDD only occur at higher levels of CUD risk (i.e. post-threshold). Additionally, it is likely that there are shared aetiological factors between CUD and MDD. Evidence from twin (Fu et al., 2002; Lynskey et al., 2004) and molecular genetic studies (Bobadilla, Vaske, & Asberg, 2013; Hodgson et al., 2016; Sherva et al., 2016) suggests that there are genetic factors influencing both cannabis involvement and MDD. There is also a plethora of environmental factors that act as risk factors for both (see Chapter 4, e.g. Fergusson and Horwood 1997 and Feingold et al. 2014). Overall, the almost identical fit of both models may indicate that there are threshold-dependent causal links between CUD to MDD which occur at the level of liability.

This interpretation is compatible with several findings. Risk factors for CUD, such as heavy cannabis use, are likely to exert an environmental and genetic effect on MDD. Heavy cannabis use can alter various domains of cognitive functioning, such as attention and memory (Solowij, 2002), and thereby affect daily functioning and potentially create circumstances in which individuals are more likely to develop MDD. For instance, cannabis use impacts educational attainment negatively (Lynskey & Hall, 2000), which in turn may affect emotional wellbeing. Environmental effects may also manifest themselves through changes in brain structure and function. Heavy cannabis users show a decrease in amygdala volumes (Yucel et al., 2008), which is also the case in unmedicated patients with MDD (Hamilton, Siemer, & Gotlib, 2008). Furthermore, cannabis intake influences the secretion of cortisol in the HPA axis and may lead to the dysregulation of such secretion over time (see Patel et al. 2014 for a

review). HPA dysregulation and an increase in cortisol levels, is also thought to play a critical role in the aetiology of MDD (Holsboer, 2000). Genes may modulate these environmental influences. Lastly, the conclusion that causal processes may be at work in individuals at high risk for CUD (e.g. high levels of cannabis use), also fits well with longitudinal studies which show that high levels of cannabis use are more strongly associated with MDD than lower levels.

6.4.2 Heritability estimates

These estimates are in line with estimates from previous analyses in this thesis (Chapter 5), similar to other twin studies for MDD (Kendler, Gatz, et al., 2006; Sullivan et al., 2000), and to studies on Cannabis Abuse and Dependence that included similar samples. Kendler et al. (2006a) report a heritability estimate of 77% (95% CI = 46%–93%) for DSM-IV Cannabis Abuse and Dependence in a sample of same-sex and opposite-sex twins with a mean age of 28.2. While a meta-analysis on twin studies reporting at least 1 symptom of Abuse or Dependence, presents lower heritability estimates (males: 54.4% (95% CI = 37.9%–64.9%), females: 58.5 (95% CI = 44.2%–72.9%), Verweij et al. 2010), the higher estimate obtained in the current study may be related to differences in sampling or the definition of problematic cannabis use.

6.4.3 Strengths and limitations

The primary strength of the current analysis is the comprehensive comparison between different models of co-morbidity. Many of these models are not commonly reported in twin studies. The best fitting model in the current analyses, the Random Multiformity model, is not a model which is usually tested in the twin literature, although it is particularly relevant to the investigation of drug use and use disorders. As an example, this model, when compared to the causal model, can test whether the likelihood of a co-morbid condition (e.g. MDD) increases at any or only at high levels of use. This is a relevant question to investigate for public health and policy.

However, difficulty in differentiating between models was a known limitation, based on previous studies. Rhee et al. (2004) demonstrated that the NK approach to

discriminating between different models of co-morbidity is valid, but they did so with a large simulated sample and still noted several challenges. They highlighted that it is particularly difficult to discriminate between the multiformity and the Correlated Liabilities model classes, which was also the case in the current sample. Additionally, Rhee et al. (2004) pointed out that discrimination within subclasses of models (e.g. RM vs. RM of CUD) is also problematic. In the current analyses, the difference within subclasses was often not more than 3 AIC. It may be beneficial to replicate the study with larger samples or use meta-analysis to examine whether differences between models become more distinct. Replication of the results in this chapter would be useful to explore whether the results are cohort-specific or generalise across cohorts, but this is outside the scope of the current thesis.

One limitation of the current study is that sex differences have not been taken into account. The prevalence of MDD and CUD did differ between males and females in the analysis sample, but currently all co-morbidity models can only be fitted on contingency tables, in which it was not possible to specify separate thresholds for males and females. The alternative approach of fitting separate models for males and females was not feasible due to reduction in sample size and associated loss of power. However, there are currently no grounds to assume that different co-morbidity models would explain co-morbid cases in males and females. For instance, Agrawal et al. (2009) examined the co-morbidity between cannabis and tobacco use, and fitted separate models for male and female twins. They found that model fits were very similar for both sexes. It may be an interesting avenue for future research to explore sex differences in larger samples or using meta-analysis.

Given that one of the best-fitting models makes assumptions about causality, it is also an important limitation that the data are retrospective and age of onset was not considered in the analyses. Using retrospective data has several disadvantages (see Section 7.3 and e.g. Coughlin 1990), but for the current analyses the most pertinent drawback is that longitudinal data would be better suited to test the direction of causation. Beyond twin models, recent molecular genetic methods also offer an interesting avenue to assess causality (see Pickrell et al. 2016).

6.4.4 Conclusion

Overall, the model fitting approach was helpful in indicating the likely relationship between CUD and MDD. While it was not possible to statistically differentiate between the two best fitting models, the RM of CUD and CUD causes MDD models, they both indicate that the direction of influence goes from CUD to MDD. Combined, the models suggest that CUD risk factors (liability) may cause MDD symptoms, but only in higher risk individuals. In addition, several models can be statistically excluded: CUD and MDD are not likely to be co-morbid by chance, arise from the same risk factors, or be due to a liability separate from the pure form of the disorders. The fact that a random multiformity model is the best fitting model is remarkable, because this model is not widely reported. Replications on larger samples would be beneficial in order to help differentiate between models with subtle differences.

7 Discussion

This thesis set out to test the presence of a co-morbidity between Cannabis Use Disorder (CUD) and Major Depressive Disorder (MDD) using epidemiological analyses, estimate the influence of genetic and environmental factors on these phenotypes and their covariance in bivariate twin model analyses, investigate a possible causal association and 13 different models of co-morbidity using the Neale and Kendler (1995) approach. The epidemiological analyses found that MDD and CUD were significantly co-morbid in this sample of 3824 Australian twins and their non-twin siblings (OR = 2.23, 95% CI = 1.84–2.70), and that this co-morbidity could not be fully attributed to various psychiatric, trauma-related, parental, peer and demographic covariates. Bivariate twin analyses found that – when separated into additive genetic, shared environmental and non-shared environmental correlations – the only significant correlation between MDD and CUD was genetic (*r* =.41, 95% CI = .24–.60). A possible causal relationship could not be excluded, because MDD and CUD were significantly associated (OR = 2.83, 95% CI = 1.12–7.19) in MZ twins discordant for both disorders. The Neale and Kendler (1995) co-morbidity analyses indicated that the direction of influence was from CUD to MDD, and that CUD risk factors may cause MDD symptoms, particularly in individuals at high risk of CUD.

This chapter briefly summarises key findings, and then discusses: a) how the findings fit into the wider theoretical framework, b) how they fit into a biological framework, c) how the findings may direct further research within and across methodologies, and d) how they could inform policy, treatment and prevention. General methodological strengths and limitations will also be discussed.

7.1 Epidemiological analyses

Epidemiological analyses involved running multivariable logistic regressions between MDD, CUD and associated covariates on the full sample of 3824 twins and non-twin siblings. The analyses had two main aims: a) to establish whether and to what extent CUD and MDD were co-morbid in this sample, and b) whether and to what extent covariates influenced this association. Covariates were chosen from an extensive

range of available cannabis-related measures and selected based on previous cross-sectional and longitudinal studies examining clinical levels of cannabis involvement and depression (see Table 2).

The main finding from these analyses was that the odds of meeting the criteria for one disorder increased significantly when an individual met the diagnostic criteria for the other (OR = 2.23, 95% CI = 1.84 – 2.70). This association remained significant when multiple covariates, including other drug dependence and psychopathology, were included in the regression analyses (aOR (MDD as outcome) = 1.96, 95% CI = 1.57–2.45; aOR (CUD as outcome) = 1.92, 95% CI = 1.53–2.41). No covariate significantly attenuated the association, although several factors were significant predictors of both MDD and CUD. These findings suggested that there may be a unique aetiological link between CUD and MDD which cannot be fully explained by covariates. However, this interpretation is tentative, as a cohort study cannot control for all possible covariates.

The epidemiological findings in this study fall within the range of results found in cross-sectional (see e.g. Degenhardt, Hall, Lynskey, Cofey, & Patton, 2012), longitudinal (e.g. Pacek, Martins, & Crum, 2013) and twin studies (Agrawal et al., 2017), which have been reviewed in Chapter 1. The previous literature had rarely examined the relationship between MDD and CUD specifically. In addition, longitudinal studies which did report an association between MDD and CUD also controlled for a small number of covariates. The epidemiological results are a valuable addition to previous literature because they demonstrate a relationship between MDD and CUD despite having controlled for a large number of covariates found to be associated with both phenotypes in previous studies.

The analyses also highlight the range of covariates which should be included in any further analyses of the association between the two main phenotypes. Of particular importance is the inclusion of conduct disorder (aOR_{MDD} = 1.55, 95 % CI = 1.13–2.12; aOR_{CUD} = 3.44, 95 % CI = 2.52–4.69), childhood sexual abuse (aOR_{MDD} = 1.73, 95 % CI = 1.30–2.30; aOR_{CUD} = 1.79, 95 % CI = 1.24–2.60) and parent-child disagreements (aOR_{MDD} = 1.88, 95 % CI = 1.45–2.44; aOR_{CUD} = 1.72, 95 % CI = 1.24–2.38). These covariates significantly increased the risk of both MDD and CUD

after the inclusion of all other covariates, and may be indicators for possible preventative measures for the co-morbidity (see Sections **Error! Reference source not found.** below, and 4.4.2 in Chapter 4).

Additionally, CUD showed significant positive associations with nicotine (aOR = 2.50, 95 % CI = 2.00–3.12) and illicit drug dependence (aOR = 1.89, 95 % CI = 1.01–3.56), as well as peer drug use (aOR = 2.19, 95 % CI = 1.48–3.24). Positive associations with MDD were found for low family SES (aOR = 1.44, 95 % CI = 1.15–1.80), panic disorder (aOR = 3.12, 95 % CI = 1.71–5.67), social phobia (aOR = 2.87, 95 % CI = 2.32–3.54) and parental problems (aOR = 2.62, 95 % CI = 1.88–3.67). These variables should also likely be included in any analysis of the co-morbidity of CUD and MDD.

7.2 Twin model findings

7.2.1 Heritability of, and correlation between, CUD and MDD

Since the co-morbidity between CUD and MDD was demonstrated to be significant in this sample, bivariate correlated liabilities models were fitted to assess the relative importance of A, C/D and E in explaining the variance of and covariance between CUD and MDD. The models were fitted using a sample of 3326 twins, and potential quantitative and qualitative sex differences were tested.

Findings from these analyses replicated those of previous studies in finding a significant heritably component of CUD (h^2 = .77, 95% CI = .67–.85) and MDD (h^2 = .42, 95% CI = .30–.54), as well as a significant genetic correlation between the phenotypes (r_g = .41, 95% CI = .24–.60). Sex differences were only found to affect disorder prevalence, i.e. the thresholds on the liability distributions. The heritability estimates (Sullivan et al., 2000; Verweij et al., 2010), genetic correlation (Lynskey et al., 2004), and sex differences in prevalence (Kessler et al., 2005, 1994) were compatible with previous studies.

The primary finding of interest is the genetic correlation, since it points to the most likely source of co-morbidity between CUD and MDD, which, according to the current findings, is genetic, since there were no other significant environmental correlations.

Although only one previous twin study has estimated a similar genetic correlation between these phenotypes (Lynskey et al., 2004), the current findings are also congruent with recent molecular genetic studies analysing common genetic variation. Among non-twin methods, genetic correlation estimates from molecular genetic studies are the closest comparison to those of twin studies.

Large scale studies have demonstrated that common genetic variation can explain a small, but significant proportion of variance in MDD (Wray & Sullivan, 2017) and CUD (Demontis et al., 2018), and that there is a significant common genetic overlap between these disorders (Carey et al., 2016; Demontis et al., 2018; Sherva et al., 2016). Two independent studies have reported a significant association between MDD and CUD polygenic risk scores (Carey et al., 2016; Demontis et al., 2018). Another study found significant pleiotropy between cannabis dependence and MDD by comparing genome-wide association study (GWAS) summary statistics for the two outcomes (Sherva et al., 2016), i.e. the same single nucleotide polymorphisms (SNPs) have been associated with both disorders. Significant pleiotropy has also been found for MDD and cannabis use (Hodgson et al., 2016).

The overall implication of the current finding is that research into the aetiology of this co-morbidity should include the consideration of genetically influenced factors. A recent EMCDDA report on the co-morbidity of substance abuse and other psychiatric disorders concludes that further research is necessary to identify individuals at high risk of developing co-morbidities between substance use and other psychiatric disorders in order to administer early interventions. The current results suggest that these risk factors are likely to be genetically influenced.

This may suggest that genetically influenced covariates, such as conduct disorder, which occur before the onset of either CUD or MDD should be targeted for preventative measures. Additionally, - since both CUD and MDD are polygenic traits (see Chapter 3) - molecular genetic studies may be useful in identifying early genetic

risk. These will likely need to identify a large number of loci, which will depend on power and consequently large sample sizes. Since familial factors play a significant role in the co-morbidity, the current results also support the use of family history in predicting patient prognosis, which was previously suggested by Milne et al. (2009). Funding research which enables genomic consortia to conduct large-scale molecular studies into the genetic risks of CUD and MDD and identifies genetically influenced early risk factors is likely to contribute toward the EMCDDA's goal of identifying high risk individuals more efficiently.

7.2.2 Causality and Neale and Kendler (1995) models of CUD and MDD co-morbidity

Although the thesis results and studies mentioned above have reported a genetic overlap between CUD and MDD, correlated liabilities may not be the best or only model of the aetiological mechanisms linking CUD and MDD. As demonstrated by discordant twin analyses, causality could not be discounted as an explanation for the co-morbidity, and 13 different models of co-morbidity (Neale & Kendler, 1995) were fitted to a sample of 565 complete MZ and 640 complete DZ twin pairs.

The key finding of these analyses was that there may be a causal effect of CUD on MDD. This finding is suggested because the CUD causes MDD model was one of the best fitting co-morbidity models and all significant models suggested a direction of effect from CUD to MDD. In combination with the second model demonstrating the best-fit to the data, the random multiformity of CUD model, the co-morbidity model findings suggest that CUD risk factors may cause MDD symptoms, but only in higher risk individuals.

Another finding which supports a possible causal relationship is the conducted discordant twin analysis. MZ twins who met the criteria for CUD were significantly more likely to meet the criteria for MDD than their co-twin discordant for CUD (OR = 2.83, 95% CI = 1.12–7.19). While this does not *prove* a causal relationship, discordant twin analyses control for a large number of possible confounding influences since twins are genetically identical and have grown up in partially overlapping environments. A lack of association between CUD and MDD in MZ twins

could be interpreted as indicating that any previously found association was due to confounding genetic or environmental variables.

These findings are in agreement with a systematic review of several longitudinal studies, which has found a significant effect of heavy (high-risk), but not light (low-risk) cannabis use on later symptoms of depression (Lev-Ran et al., 2014). Other methods of validation are currently lacking and will be suggested in the following sections.

Although the results presented in this study cannot definitively answer whether the link between CUD and MDD is causal, it provides critical evidence in favour of a causal association, and demonstrates that genetic designs can be an important tool in distinguishing between different types of possible co-morbidities. This should prompt further investigation, which goes beyond longitudinal research designs and capitalises on the benefits of genetically informed designs.

It is unclear why longitudinal studies have found mixed evidence for the causal link between CUD and MDD while a causal model provides the best fit in the current twin design. One possibility is that there may be multiple forms in which CUD and MDD can co-occur. Having found different, equally fitting models may be either due to power constraints, or it may be an important finding in its own right. The different models which provide a similarly good fit to the data in the Neale and Kendler (1995) approach may do so because of multiple aetiological pathways to the co-morbidity of these disorders.

Additionally, the current findings suggest that in trying to identify risk factors for the co-morbidity between CUD and MDD, researchers may want to focus on genetic risk factors primarily associated with CUD. The good fit of the random multiformity of CUD model supports the idea that research should focus on high-risk individuals, because the model suggests that only individuals at high risk of developing CUD will develop MDD. Since CUD was found to be primarily influenced by genetic factors, as explained in Chapter 5, factors placing individuals at high risk of CUD are likely to be genetic.

7.3 General methodological strengths and limitations

7.3.1 Strengths

A primary strength of this thesis is the application of multiple methodologies to address different core questions around the covariates, source and process underlying the co-morbidity between MDD and CUD. Epidemiological analyses capitalised on analysing the full sample, including 476 non-twin siblings, and tested the influence of a multitude of covariates. Bivariate correlated twin models were applied to decompose the covariance between MDD and CUD into genetic and environmental factors to get a clear understanding of the source of co-morbidity, while discordant and co-morbidity model analyses investigated the processes which might underlie co-morbidity. Answering questions of epidemiological and behavioural genetic nature is a unique advantage of twin samples and of the analyses in this thesis.

Twin studies are currently the most comprehensive method to compare between a large number of different co-morbidity models at the same time. They are also a powerful method to estimate genetic correlations between traits, and currently cannot be replaced by molecular genetic studies due to the issue of missing heritability (Manolio et al., 2009). Presently it is still not possible for molecular genetic studies to account for all the genetic effects that can be detected in twin studies (Nolte et al., 2017).

A further strength of the studies presented here lies in the size and nature of the sample analysed. Data collection was tailored to the aim of analysing cannabis and related variables (Lynskey et al., 2012). Consequently, the sample contains a large number of cannabis-related measures and covariates, which is an advantage compared to other cohort studies, which may analyse cannabis-related variables in datasets originally intended for a different purpose. For instance, the NESARC (National Epidemiologic Survey on Alcohol and Related Conditions) is an epidemiological survey centred around alcohol use and related problems, although many longitudinal studies (Blanco et al., 2016; Cougle et al., 2015; Feingold et al.,

2015; Pacek et al., 2013) on cannabis-related issues have been published using this data.

The number and quality of measurement of available covariates has been of critical benefit to the epidemiological analyses, which were able to include a wide variety of covariates compared to some other cohort studies examining the relationship between CUD and MDD (see Table 2). Since there is some inconsistency in the literature about the strength and significance of the associations found between CUD and MDD or other clinical levels of cannabis involvement and depressive symptoms, adequate control for covariates is crucial.

This sample also maximises available power given its size. It contains relatively small numbers of missing data, since computer-assisted telephone interviews were used. Additionally, a large proportion of individuals met criteria for CUD (similar to cannabis dependence estimates in a previous Australian cohort; Lynskey et al., 2004), which increased the power of the analyses. Twin studies with phenotypes which have low population prevalence suffer from power issues (Neale, Eaves, & Kendler, 1994) and difficulties distinguishing between sources of variation (Neale & Maes, 2004).

7.3.2 Limitations

General limitations of twin studies have been discussed in Section 3.7 (Chapter 3). Findings from twin studies may be affected by a violation of the Equal Environments Assumption, assortative mating and gene-environment or gene-gene interplay. As reviewed in Section 3.7, violations of the Equal Environments Assumption have not previously been found for MDD or cannabis dependence (Kendler et al., 1993, 1994; Lynskey et al., 2002; Xian et al., 2000). Additionally, assortative mating is unlikely to bias twin models fit to psychiatric phenotypes. (Maes et al., 1998). The biasing effect of gene-environment and gene-gene interplay on A, C and E estimates is less clear. Since the dataset does not contain molecular genetic information, they were not modelled and are beyond the scope of this thesis.

Data collection relied on retrospective self-report measures, which is a further potential limitation. Since cannabis is an illegal substance in Australia, individuals may not accurately self-report patterns of drug use and use disorders. However, the relatively high prevalence of cannabis use (68.52%) suggests that under-reporting of drug use is unlikely to be problematic in this sample. Additionally, specific experiences (e.g. age of first symptom onset) are unlikely to be directly observed, so there is no viable alternative to self-reports (Wagner & Anthony, 2002). The retrospective nature of reporting and consequent recall bias is a concern for epidemiological analyses in particular, because these relied on age of onset reports on a number of variables. One could argue that cannabis-related behaviours may not be remembered correctly due to effects on memory (Hall, 2015), but studies suggest that cannabis-related experiences are recalled reliably (Johnson & Mott, 2001). Despite potential biases, self-report retrospective studies have been used widely to examine both depression and cannabis involvement (see studies reviewed in Chapter 1) and are considered a valid method of data collection in general and in drug users (Anglin, Hser, & Chou, 1993; Sartor et al., 2011).

Additionally, the current data does not permit the examination of whether the association between CUD and MDD varies by type or potency of cannabis used. Since cannabis potency may influence the likelihood of developing cannabis dependence (Freeman & Winstock, 2015), as well as other psychiatric disorders, information on type of cannabis used would have clarified, for instance, whether the causal link suggested in the co-morbidity models may have been driven by a sub-population of high potency cannabis users. Although at time of writing the author is not aware of any studies that have examined the effect of cannabis type or potency on MDD, some research suggests that the likelihood of schizophrenia following cannabis use depends on the amount of THC and relative proportion of THC and cannabidiol (CBD), two phytocannabinoids, in the cannabis smoked. For example, one study found that cannabis with high levels of THC and an large ratio of THC to CBD was three to five times more frequent in individuals with first episode psychosis (Di Forti et al., 2015). The lack of potency and cannabis type information is a limitation of the current study and an interesting avenue for future research.

7.4 Synthesised analysis of thesis findings

A synthesised view of the findings might suggest that CUD and MDD are significantly co-morbid, and that this co-morbidity is likely to be explained through a causal effect of CUD on MDD which may be influenced by genetic factors. The following section will attempt to form a coherent picture from the findings that: a) CUD and MDD are significantly co-morbid, b) the only significant correlation between CUD and MDD was found to be genetic, c) the direction of effect in the co-morbidity models is from CUD to MDD, and d) that MZ twins who met criteria for CUD were significantly more likely to meet criteria for MDD than their co-twins who did not meet the criteria for CUD. This will be done by fitting them into a theoretical and biological framework.

7.4.1 Integration of findings into a theoretical framework

In an overview article Agrawal and Lynskey (2014) have suggested several potential mechanisms by which CUD-related factors may influence the development of MDD (see Figure 14). The findings in this thesis most support the models in Figure 14b and Figure 14e. Due to the significant genetic correlation in Chapter 5, with no significant shared or non-shared environmental correlation, the current findings add evidence to the suggestion that any aetiological mechanism is mediated through genetically-influenced factors. Due to the significantly better fit of models suggesting that the direction of effect goes from CUD to MDD, the current findings suggest that a causal model may be a more suitable aetiological framework than a model which implies correlated liabilities only. A causal genetically-mediated model may take two forms according to Agrawal and Lynskey (2014). CUD use may impact genetically-influenced biological structures, such as limbic brain areas and the hypothalamic–pituitary-adrenal axis, and the susceptibility to its effect on MDD may be genetically influenced (Figure 14b). Alternatively, smoking large quantities of cannabis due to CUD may lead to an *activation of diathesis*, i.e. an activation of genetic factors which increase the likelihood of developing MDD (Figure 14e). If it were the case that shared genes simultaneously influence CUD and MDD, one would have expected the correlated liabilities model (Figure 14d) to be the best fit in the Neale and Kendler (1995) models. However, it was a significantly worse fit than models suggesting a causal relationship.

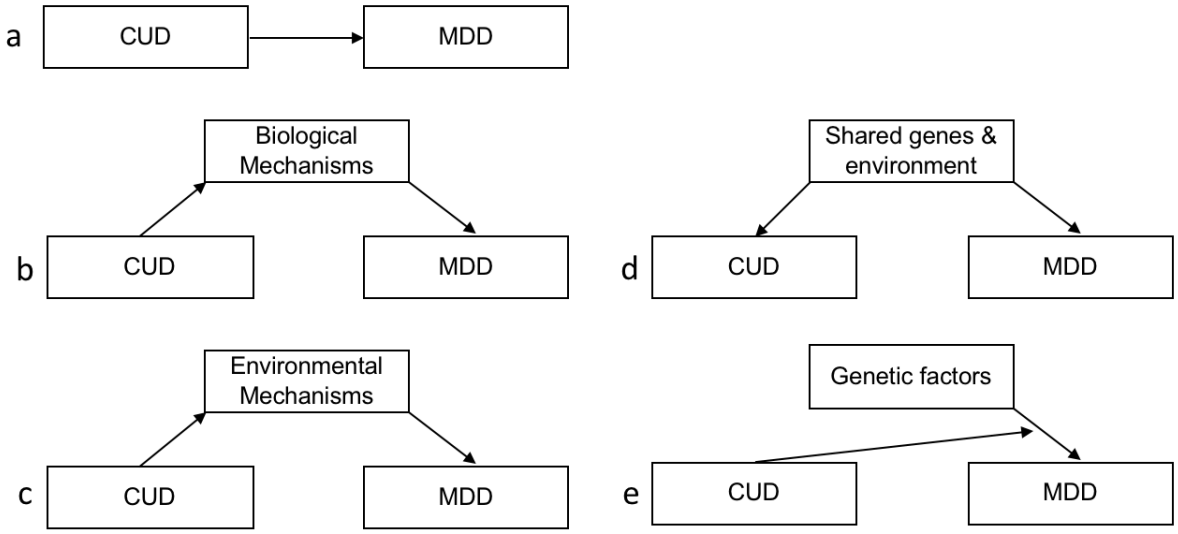

Figure 14. Possible aetiological relationships between CUD and MDD, adapted from Agrawal and Lynskey (2014).

Evidence in favour of the activation of diathesis (e) model comes not only from the suggestion of causal genetic influences, but also from a significantly elevated rate of MDD among MZ twins who had CUD compared to their discordant co-twins without CUD. Both MZ twins have the same genetic material, but the exposure of one twin to high levels of cannabis may activate genes that increase the likelihood of MDD. This may explain why a causal mechanism cannot be excluded based on discordant twin analyses, but an environmental link between CUD and MDD is not evident in the bivariate twin analyses. The discordant twin results could also be consistent with model (b), since the genetic mechanisms impacting biological structures altered by cannabis use and producing MDD may only become relevant when high levels of cannabis use are present.

For either model, the section below aims to give a thorough overview of potential biological 'bridges' between CUD and MDD which may be the location of any genetic effects.

7.4.2 Integration of findings into a biological framework

The most likely biological bridge between CUD and MDD involves the endocannabinoid system, and its influence on the limbic system, as well as the

hypothalamic–pituitary–adrenal axis. High levels of cannabis use are likely to modify signalling in both systems, which in turn may lead to symptoms of depression.

7.4.2.1 Endocannabinoid influence on the limbic system and HPA axis

The limbic system is crucially involved in emotional processing, contains moderate to high amounts of endocannabinoid receptors (Hill et al., 2008), and functional disturbances of the limbic system are linked to symptoms of MDD (Herman, Ostrander, Mueller, & Figueiredo, 2005). In a comprehensive summary on the link between the endocannabinoid system and depression, Hill et al. (2008) review evidence showing that levels of endogenous cannabinoids are altered in individuals with MDD, as is the expression of CB_1 receptors. Endocannabinoid receptor signalling is also altered in the course of various treatments for MDD.

Critical evidence suggesting a relationship between endocannabinoid functioning and MDD comes from human trials on the anti-obesity drug rimonabant, which is a partial CB1 receptor agonist. A meta-analysis of randomised controlled trials concluded that that the drug likely caused MDD, depressive mood and depressive symptoms (Christensen, Kristensen, Bartels, Bliddal, & Astrup, 2007). A plausible interpretation of these results is that MDD may have been brought on by the reduced ability of endocannabinoids to bind to CB1 receptors. Since CB1 receptors were found to be down-regulated in cannabis users until they were abstinent for a month (Hirvonen et al., 2012), the effect of cannabis use on CB1 receptors is a plausible neurochemical explanation for the association between CUD and MDD.

In addition, specific functions of the endocannabinoid systems within the limbic system can be linked to specific symptoms of MDD. Hill et al. (2008) examined changes in endocannabinoid levels and CB_1 receptor density in response to chronic unpredictable stress in rats – an animal model of depression. They found that CB_1 receptor density was altered in most regions of the limbic system and that levels of anandamide were significantly reduced in all regions. The endocannabinoid system helps to maintain reward salience via its activity in the limbic brain regions (Parsons & Hurd, 2015). Anhedonia is a main symptom of MDD and is characterised by a lack

of response to things and activities previously perceived as rewarding. This may be due to abnormalities in endocannabinoid signalling.

However, the key influence of the endocannabinoid system on depression may extend beyond the limbic regions, via the hypothalamic–pituitary–adrenal (HPA) axis. The human stress response is regulated by a cascade of chemical changes involving the hypothalamus, pituitary gland and adrenal cortex. When people are subjected to a stressor, the hypothalamus secretes vasopressin and adrenocorticotropic hormone-releasing factor, which triggers the release of adrenocorticotropic hormone in the pituitary. When the latter hormone reaches the adrenal cortex, it causes the release of adrenal steroids, such as glucocorticoids, which help generate an appropriate stress response. Glucocorticoids suppress the immune system and activate the metabolism, preparing the body to address the stressor.

Glucocorticoids also directly impact the brain. Particularly relevant for depression, glucocorticoids affect the functioning of limbic regions, such as the hippocampus and amygdala, and interact with monoamines, serotonin and noradrenaline. Excess amounts of glucocorticoids make hippocampus cells more susceptible to damage and decrease their proliferation (Herbert et al., 2006). Furthermore, glucocorticoids interact with noradrenaline in the amygdala to enhance memory formation in response to emotionally arousing stimuli (Herbert et al., 2006). Glucocorticoids are likely to play an important role in adaptive memory formation to improve an individual's response to aversive events, but may also play a role in the development of depression by increasing attention or memory toward negative events (Kukolja et al., 2008). Glucocorticoids also regulate tryptophan hydroxylase, an enzyme required for the synthesis of serotonin, as well as the expression of serotonin receptors (Herbert et al., 2006), and serotonin dysregulation is crucially involved in the pathophysiology of depression (Ressler & Nemeroff, 2000).

It is not surprising that one of the most consistent biological correlates of MDD is a hyperactivity of the HPA axis (Pariante, 2017; Pariante & Lightman, 2008). Individuals with MDD have increased adrenal and pituitary gland activity, and higher levels of cortisol in various bodily fluids (Pariante & Lightman, 2008). Additionally, individuals treated with synthetic glucocorticoids are at increased risk of developing

depressive symptoms (Brown & Suppes, 1998). HPA hyperactivity is thought to be a cause, rather than consequence of MDD, and result from impaired negative feedback of glucocorticoids on the stress response cascade (Pariante & Lightman, 2008).

Since endocannabinoid levels change when people are subjected to stress (Hill, Miller, Carrier, Gorzalka, & Hillard, 2009), endocannabinoids may be involved in the modulation of the HPA axis. Several studies on mice suggest that endocannabinoids serve to inhibit HPA axis activation (Gorzalka et al., 2008). Endocannabinoids are also thought to play a role in stress adaptation, as their increase during chronic stress dampens the stress response (Gorzalka et al., 2008). This is a crucial function, as steroid release is useful for combating stressors in the short term, but causes harm in the case of overexposure, as reviewed above.

However, cannabis intake influences the secretion of cortisol in the HPA axis and may lead to the dysregulation of such secretion over time (see Patel et al. 2014 for a review). If endocannabinoids fail to successfully dampen the stress response, this may modulate the effect of stress on depression. One molecular genetic study has shown that polymorphisms in the CNR1 gene, which encodes the CB1 receptor, moderated the influence of childhood physical abuse on anhedonia (Agrawal et al., 2012). In other words, genetic factors primarily associated with CUD may contribute to the development of MDD through their effect on the stress response.

The evidence reviewed above points to a link between the endocannabinoid system, CUD and MDD. Genetic factors may influence neurobiological structure and function predisposing an individual to both MDD and CUD or providing a link for causal influences.

7.5 Implications for future research

7.5.1 Multivariable twin models

Since epidemiological analyses found a number of covariates which may have attenuated the association between CUD and MDD, albeit not significantly, and

bivariate twin modelling found a significant genetic correlation between the two phenotypes, it would be a valuable next step to examine the influence of these covariates on the genetic correlation.

For instance, one could examine whether the genetic correlation between CUD and MDD – which was observed in Chapter 5 – is cannabis-specific, or shared between different substances. Studies from a range of fields have observed that liability is likely to be shared between different drug phenotypes to a significant extent (Moss, Chen, & Yi, 2014; Palmer et al., 2015, 2012; Schwantes-An et al., 2016; Vanyukov et al., 2012). A significant attenuation of the genetic correlation between CUD and MDD after the inclusion of a third substance would suggest that the genetic correlation between CUD and MDD is not cannabis-specific. Since poly drug use occurs frequently (Moss et al., 2014) and significant genetic links between MDD and other substance use have been found in the literature (Carey et al., 2016; Edwards & Kendler, 2012; Kendler, Neale, MacLean, et al., 1993), the specificity of the correlation to cannabis should be examined before any health or policy decisions are made on the basis of the link between CUD and MDD.

Additionally, it may be valuable to examine whether the genetic correlation is MDD-specific or shared with multiple forms of non-substance psychopathology. A molecular genetic study employing polygenic risk score analyses found that there was a significant association between polygenic risk scores for a substance-involvement cluster (representing a general liability to alcohol, cannabis, cocaine, nicotine, and opioid involvement) and a psychiatric disorder cluster (MDD, ADHD, autism spectrum disorder, bipolar disorder and schizophrenia), suggesting that multiple forms of psychopathology share genetic influences (Carey et al., 2016). This finding has been supported by polygenic risk score association analyses of the Cross-Disorder Group of the Psychiatric Genomics Consortium (2013) for the aforementioned five psychiatric disorders.

To ascertain the specificity of the genetic link found in Chapter 5, multivariate analyses could be conducted using twin models or molecular genetic methods. However, both would depend on sample constraints and multivariate twin models may prove difficult to fit with clinical phenotypes, as binary or ordinal multivariate

data often causes multi-parameter optimisation problems in OpenMx (Neale et al., 2016). The current sample may therefore not be suitable and larger samples would be best placed to examine this question.

7.5.2 Alternative causal models

Although a large number of aetiological models can be compared using cross-sectional twin data, they are not the gold-standard test of causality. A randomised controlled trial would be unethical, but there are several other ways to improve on the design of the current study to estimate causality.

7.5.2.1 Longitudinal twin research

Firstly, it could be advantageous to examine the co-morbidity models in a longitudinal twin sample. Both MDD and CUD would have to be measured at baseline and at follow-up, allowing a comparison between co-morbidity models which make predictions in both directions of causality. This may also improve the degree to which co-morbidity models can be differentiated from each other. To the best of the author's knowledge, no longitudinal twin study has examined co-morbidity at the time of writing.

7.5.2.2 Mendelian randomisation

Secondly, there are novel molecular genetic methods that may be able to circumvent the ethical problems associated with randomised controlled trials. Mendelian randomisation is one approach that could validate the possibly of causal genetic relationships between CUD and MDD reported in this thesis. Instead of grouping participants into individuals who do and do not consume large amounts of cannabis, individuals at genetic risk of developing CUD are treated as the 'affected' and individuals at low risk as the 'control' group (Lawlor, Harbord, Sterne, Timpson, & Davey Smith, 2008). Genetic risk can be operationalised as a polygenic risk score for CUD for each individual. It can then be analysed whether individuals at genetic risk for CUD are significantly more likely to have MDD. This type of analysis does not depend on longitudinal data, because any measured genetic risk would precede the

development of a disorder (Davey Smith & Hemani, 2014). Furthermore, Mendelian Randomisation can now be combined with twin models to circumvent problems involving pleiotropy (Minica, Dolan, Boomsma, de Geus, & Neale, 2017).

However, this method is dependent on having strong genetic predictors of the outcomes in question (Burgess & Thompson, 2011), and although large-scale GWAS of CUD and MDD have been performed, the proportion of variance that can be explained by identified associated loci is low. Nonetheless, this would be an interesting area to explore as genetic loci explaining a larger proportion of variance in CUD and MDD are discovered.

7.5.2.3 Genetic pathway analysis

Thirdly, given the likely biological pathways of overlap between CUD and MDD and the fact that their relationship was suggested to be primarily genetic in bivariate analyses, an analysis of specific genes and pathways underlying the genetic correlation between these outcomes could aid in elucidating the causal relationship between CUD and MDD. The current study cannot provide information on the specific sources (e.g. single genes) of their genetic correlation or causation.

One approach for this is the systematic comparison of known gene or pathway-associations for each outcome. One of the candidate genes which may be associated with both CUD and MDD is CNR1, as mentioned above (Agrawal et al., 2012). Additionally, a recent study has tested the association between substance use, depression and genetic variants with four candidate genes: the serotonin transporter (5-HTTLPR) gene, the neuropeptide Y (NPY) gene, the brain-derived neurotropic factor (BNDF) gene and the corticotrophin-releasing hormone-binding protein (CRHBP) gene (Trucco, Villafuerte, Hussong, Burmeister, & Zucker, 2018). The three latter genetic variants are all associated with the human stress response. Trucco et al. (2018) examined a sample of 426 adolescents and analysed whether variants of the aforementioned genes were linked to adolescent substance use and mediated by childhood depressive symptoms. Although none of the associations examined in this study were significant after covariates were controlled for, perhaps due to the small sample size and the fact that variables were measured in

adolescence and included substance use, rather than substance use disorders, their approach is promising in establishing biologically-informed pathways involving internalising psychopathology, such as MDD, and substance use disorders, such as CUD.

An alternative approach for elucidating the genetic overlap between CUD and MDD is the hypothesis-free comparison of GWAS summary statistics for both phenotypes. This could be achieved by looking for concordance in loci showing suggestive ($1x10^{-5}$) or genome-wide ($5x10^{-8}$) significance between outcomes or comparing gene- or pathway-based association results. A similar but perhaps more automated approach to highlight genes or pathways underlying the genetic correlation between CUD and MDD is through the estimation of partitioned genetic correlation, whereby the genetic correlation is estimated using subsets of genetic variation (Ni, Moser, Wray, & Lee, 2017).

7.5.2.4 Ecological Momentary Assessment

Additionally, causality between CUD and MDD may occur over shorter time windows than those that are usually investigated in longitudinal studies. Twin, longitudinal and cross-sectional studies often rely on the reporting accuracy over long periods of time, which may be inconsistent, and explain the mixed results with respect to the association between CUD and MDD. To address both issues, studies utilising ecological momentary assessment (EMA), measure drug use and mood fluctuations in real time and might provide insight into moment-to-moment causality. EMA is a research design which is used to collect real-time data about participants' behaviour (e.g. cannabis use and mood fluctuations) in their natural environment (Shiffman, 2009). Several studies have looked at the association between cannabis use and mood (Buckner, Crosby, Silgado, Wonderlich, & Schmidt, 2012; Buckner, Crosby, Wonderlich, & Schmidt, 2012; Buckner et al., 2015), but none have done so in individuals who have been diagnosed with MDD and CUD.

7.6 Clinical and policy implications

7.6.1 Clinical implications

7.6.1.1 Clinical research

Although the current findings may have only limited direct clinical implications, as they are based on analyses of a general population sample, they can suggest testable hypotheses for clinical research. In particular, finding that the Random Multiformity of CUD model fit best in the co-morbidity model analyses suggests that the effect of cannabis involvement on depression should be investigated further in clinical studies. The most suitable study design to delineate whether cannabis involvement causes depressive symptoms would be a Randomised Controlled Trial. As mentioned previously, such a trial would not be ethical to conduct, although cannabis users themselves report using cannabis to treat depression (Hakkarainen et al., 2015).

Two trials conducted in the 1970s did examine a potential medicinal effect of THC on MDD, one did not report an improvement and one reported a worsening of symptoms (Turna, Patterson, & Van Ameringen, 2017). Since recent evidence suggests that cannabis products which include THC may increase the likelihood of depressive symptoms (see Chapter 1), it is clear why more recent trials have not been conducted. However, symptoms of depression could be measured as part of ongoing and future trials which measure the potential *medicinal* effect of cannabis on other physical and mental health disorders, for which beneficial effects can be expected, and ethical clearance can be obtained.

The hypothesis, based on findings from this thesis, would be that medicinal cannabis which contains THC may lead to an increase in depressive symptoms, whichever other disorder it is being used to treat. Trial results are expected to vary depending on the type of cannabis used, and trials of interest are those that include *cannabis with THC*. Many trials examine the effect of cannabidiol alone or the effect of synthetic pharmaceutical drugs (e.g. dronabinol) mimicking some properties of phytocannabinoids. However, pharmaceutical cannabinoids, cannabidiol or any

cannabis without THC are not the cannabis smoked recreationally by the general population. Consequently, results of such trials are not entirely comparable to results of population surveys examining cannabis use (and reporting an increased OR between cannabis involvement and depression), including the sample utilised in this thesis. Therefore, it is unclear whether an effect of pharmaceutical or CBD-based cannabinoids on depressive symptoms should be expected. However, the literature reviewed in Chapter 1 suggests that cannabis including THC may increase risk of depression after frequent administration.

According to a review and meta-analysis of RCTs by Whiting et al. (2015) the effect of medicinal cannabinoids, with or without THC, has been tested on nausea and vomiting, psychosis, sleep disorders, appetite stimulation in AIDS/HIV, intraocular pressure in glaucoma, spasticity due to multiple sclerosis or paraplegia, Tourette syndrome and anxiety disorders. For a variety of reasons most of these trials were assessed to be at substantial risk of bias. Interestingly, the main reason was that trials had high withdrawal rates, which were not adequately accounted for. While beyond the scope of this thesis, it would be worth investigating whether withdrawal rates may have been due to any effects cannabis may have exerted on mood.

Out of 505 studies which were assessed in the meta-analysis (Whiting et al., 2015), only five studies included a depression outcome measure in their trials, but none met criteria for inclusion in the meta-analysis. Out of the five studies, Whiting et al. (2015) report that one found negative effects of a high dose of nabiximol (containing both THC and CBD) spray on depression scores, when compared to placebo. However, the original paper does not report these results (Portenoy et al., 2012) and states that there were no significant effects of nabiximol on depression scores. None of the five studies reported results at follow-up. Another recent meta-analysis of both RCT and non-RCT studies (Goldenberg, Reid, IsHak, & Danovitch, 2017) found that three studies reported positive effects of cannabis on depressive symptoms, but only one of these was an RCT, it had a very small sample size (N = 23) and measured depression scores only during the time cannabis was administered (Ware et al., 2010), when a positive effect on mood may be expected. For a useful estimate of the effect of cannabis on depression, a follow-up measure would be necessary.

Overall, there may be considerable scope to investigate the effect of cannabis-based medicines on depression outcomes in such RCTs, because the quality of evidence is in need for improvement. It would be important to utilise a reliable and valid depression measure, preferably measuring DSM or ICD clinical symptoms, which was not used in any of the trials reviewed above. It would also be important to measure depression after a follow-up period, because the short-term effects of cannabis on mood, whether they are euphoria (Whiting et al., 2015) or depressive symptoms (e.g. Tramèr et al., 2001) cannot differentiate between temporary states and persistent disorders. A causal relationship between cannabis involvement and MDD would therefore need to be demonstrated at longer term follow-up.

As an alternative study design, a measure of depressive symptoms could be included as part of clinical trials which measure the efficacy of CUD treatments, to test whether the treatment of cannabis-related problems also decreases the likelihood of depressive symptoms (given that untreated CUD may lead to MDD). The complication is that clinical trials usually aim to recruit individuals with "pure" disorders and consequently any co-morbid symptoms of depression may be screened out at the start of a trial (e.g. Hoch et al., 2014). However, if one assumes that a portion of the individuals with CUD had "not yet" met criteria for depression but were at risk to do so in the future, it would still be informative to assess whether those who had been treated for CUD would also be less likely to later on develop depressive symptoms than those who did not.

Research into treatments for CUD and other cannabis related problems is ongoing, with no currently approved pharmacotherapies (Copeland, Pokorski, & Gibson, 2017). Among psychosocial therapies, a recent Cochrane review concluded that most support exists for a combination of Motivational Enhancement Therapy and Cognitive Behavioural Therapy (Gates, Sabioni, Copeland, Le Foll, & Gowing, 2016). Although these therapies are promising, several drawbacks have been identified. Crucially, trials lack an adequate assessment of other aspects of participant mental health (Copeland et al., 2017), which have, for instance, been implemented in studies on the efficacy of methamphetamine treatment (Mcketin et al., 2013). Addressing this drawback may provide a crucial opportunity to include reliable

measures of depressive symptoms to be used as outcomes during follow-ups of RCTs assessing therapies for CUD and other cannabis-related problems.

7.6.2 Prevention

With budget cuts to the UK's National Health Service addiction services (Mohammadi, 2014) and calls for researchers to reduce waste and increase value (e.g. Macleod et al., 2014), there is a continuous pressure to focus resources where they are needed most.

Economic analyses weighing the costs and benefits of investment into depression treatment and prevention suggest that the economic benefits will far outweigh the costs (Chisholm et al., 2016). Since prevention is preferable to treatment and treatment for co-morbid cases is known to be more challenging than for non-comorbid mental health issues (Torrens et al., 2015), it is helpful to identify high-risk groups before they develop clinical symptoms of MDD. All results in this thesis, especially co-morbidity model findings, suggest that individuals with CUD qualify as such a group. As mentioned previously, the number of European first-time treatment seekers for cannabis problems has almost doubled in the past decade (EMCDDA, 2017). These treatment seekers may be a critical target population for MDD selected preventive interventions. The goal would be to reduce the incidence of new co-morbid cases through early intervention in high risk groups, which aligns with the recommendations of the EMCDDA (Torrens et al., 2015). The most effective preventive interventions for MDD are CBT-based and interpersonal psychotherapy-based interventions (Muñoz, Cuijpers, Smit, Barrera, & Leykin, 2010). It may be crucial to aim such interventions at young individuals since the vast majority of mental health problems begin before the age of 18 (Wykes et al., 2015)

In fact, epidemiological analyses have highlighted a cluster of childhood variables - conduct disorder, childhood sexual abuse and disagreements with parents - which significantly increase the odds of both CUD and MDD. This finding may point towards the importance of childhood interventions to prevent later onset mental health co-morbidity. As previously outlined in Chapter 4, dysfunctional parenting in childhood is related to childhood sexual abuse (Fergusson et al., 1996; Putnam,

2018), conduct disorder (Furlong et al., 2013; Ogden & Hagen, 2008), as well as CUD and MDD in the current analyses. Dysfunctional parenting may therefore be a nexus worth targeting with parental training interventions in order to prevent co-morbid mental health conditions later in life.

7.6.3 Policy

Results from this thesis primarily have future research and clinical implications, because studies have not directly measured the impact of cannabis-related policy on cannabis use prevalence, CUD prevalence or cannabis-related harms. Therefore, any extrapolations from the current thesis findings to policy are tentative.

Debate around the legalisation of cannabis is ongoing and controversial, with further research needed to assess the impact of legalisation on public health (Wilkinson, Yarnell, Radhakrishnan, Ball, & D'Souza, 2016). Firstly, it has been noted in previous research that the extent of cannabis use is related to individual (Bachman, Johnston, & O'Malley, 1998) and social (Keyes et al., 2011) beliefs about its risks. In the United States, attitudes about the harmfulness of cannabis increased in some states after legalisation, but not in others (Cerdá et al., 2017). Educating the public on the possible harms of cannabis, whether legal or illegal, may be a crucial step in mitigating cannabis-related harms.

Epidemiological findings in this thesis, which are similar to findings from many previous cross-sectional studies, suggest that CUD is associated with an almost twofold increase in the likelihood of MDD., and discordant as well as co-morbidity twin analyses suggest that there may be a causal link from CUD to MDD. Given the high baseline prevalence of MDD and its disabling effects (Ferrari et al., 2013), CUD should be regarded as an important potential risk factor for MDD, since it may double an already high risk of the disorder. Although there is no conclusive proof of a causal association, the consistent evidence of a significant co-morbidity in this and other studies may suggest that educational programs in schools, medical advice to patients known to consume cannabis, and buying information in areas where cannabis is legal, should include information about this possible link and hopefully reinforce the belief that the consumption of cannabis has significant associated risks.

Although the well-researched link between cannabis use and schizophrenia could also act as strong deterrent against heavy cannabis use, schizophrenia has a lifetime prevalence of less than 1% in most populations (McGrath, Saha, Chant, & Welham, 2008). Individuals deciding whether to smoke cannabis may be more deterred by a highly prevalent disorder such as MDD, since they are more likely to know affected individuals, and a twofold increase in the risk of a prevalent disorder may inspire more caution.

Results from the co-morbidity model analyses suggest that the increased risk for MDD may occur discontinuously and only set in once high levels of CUD risk are reached. These high levels of risk may be due to high levels of cannabis use. This finding is consistent with a meta-analysis of longitudinal studies' findings that only high levels of cannabis use show a significant association with MDD (Lev-Ran et al., 2014). Given that the total prohibition of cannabis use may not be in the best public interest (MacCoun & Reuter, 2011), it will be important to advise individuals on safer levels of use and harms associated with high levels of use.

It can be argued that the disadvantages of cannabis legalisation are likely to be an increase in drug use overall, while the likely advantage would be decreased harm per use (MacCoun & Reuter, 2011). Recent trends in cannabis consumption in the US support this idea: a large scale epidemiological study has found that cannabis use, within a decade of increasing legalisation, has increased between 2001–2002 and 2012–2013, while the proportion of individuals with cannabis use disorder has decreased (Hasin et al., 2015). Given the high likelihood of an increased availability of cannabis in the future, and therefore its increased consumption, the results of this study aim to contribute to decreasing harm in users, particularly those who may be concerned about risks of developing co-occurring mental health disorders such as MDD.

7.7 Final remarks

Epidemiological and twin analyses in this thesis have provided insight into covariates, sources and processes behind the co-morbidity of MDD and CUD. While direct implications of these findings are primarily focused on future research, they

also add to the body of evidence informing decisions on policy, treatment and prevention. In summary, findings from this thesis suggest that CUD and MDD are significantly co-morbid, and associated with several childhood risk factors. The co-morbidity can be primarily attributed to genetically influenced factors and causal mechanisms, with high levels of CUD risk predisposing to the development of MDD.

8 References

Agrawal, A., & Lynskey, M. T. (2014). Cannabis controversies: How genetics can inform the study of comorbidity. *Addiction*, *109*(3), 360–370. https://doi.org/10.1111/add.12436

Agrawal, A., Lynskey, M. T., Bucholz, K. K., Martin, N. G., Madden, P. A. F., & Heath, A. C. (2007). Contrasting models of genetic co-morbidity for cannabis and other illicit drugs in adult Australian twins. *Psychological Medicine*, *37*(1), 49–60. https://doi.org/10.1017/S0033291706009287

Agrawal, A., Neale, M. C., Prescott, C. A., & Kendler, K. S. (2004). Cannabis and Other Illicit Drugs: Comorbid Use and Abuse/Dependence in Males and Females. *Behavior Genetics*, *34*(3), 217–228. https://doi.org/10.1023/B:BEGE.0000017868.07829.45

Agrawal, A., Nelson, E. C., Bucholz, K. K., Tillman, R., Grucza, R. A., Statham, D. J., … Lynskey, M. T. (2017). Major depressive disorder, suicidal thoughts and behaviours, and cannabis involvement in discordant twins: a retrospective cohort study. *The Lancet Psychiatry*, *4*(9), 706–714. https://doi.org/10.1016/S2215-0366(17)30280-8

Agrawal, A., Nelson, E. C., Littlefield, A. K., Bucholz, K. K., Degenhardt, L., Henders, A. K., … Lynskey, M. T. (2012). Cannabinoid receptor genotype moderation of the effects of childhood physical abuse on anhedonia and depression. *JAMA Psychiatry*, *69*(7), 732–740. https://doi.org/10.1001/archgenpsychiatry.2011.2273

Agrawal, A., Silberg, J. L., Lynskey, M. T., Maes, H. H., & Eaves, L. J. (2010). Mechanisms underlying the lifetime co-occurrence of tobacco and cannabis use in adolescent and young adult twins. *Drug and Alcohol Dependence*, *108*(1–2), 49–55. https://doi.org/10.1016/j.drugalcdep.2009.11.016

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, *52*(3), 317–332.

https://doi.org/10.1007/BF02294359

American Psychiatric Association. (2000). *Diagnostics and Statistical Manual of Mental Disorders* (4th ed.). Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Arlington, VA: Author.

Anglin, M. D., Hser, Y.-I., & Chou, C.-P. (1993). Reliability and Validity of Retrospective Behavioral Self-Report By Narcotics Addicts. *Evaluation Review*, *17*(1), 91–108. https://doi.org/10.1177/0193841X9301700107

Ashton, C. H., & Moore, P. B. (2011). Endocannabinoid system dysfunction in mood and related disorders. *Acta Psychiatrica Scandinavica*, *124*(4), 250–261. https://doi.org/10.1111/j.1600-0447.2011.01687.x

Bachman, J. G., Johnston, L. D., & O'Malley, P. M. (1998). Explaining recent increases in students' marijuana use: Impacts of perceived risks and disapproval, 1976 through 1996. *American Journal of Public Health*, *88*(6), 887–892. https://doi.org/10.2105/AJPH.88.6.887

Baillie, A. J., & Teesson, M. (2010). Continuous, categorical and mixture models of DSM-IV alcohol and cannabis use disorders in the Australian community. *Addiction*, *105*(7), 1246–1253. https://doi.org/10.1111/j.1360-0443.2010.02951.x

Berkson, J. (1946). Limitations of the application of fourfold table analysis to hospital data. *Biometric Bulletin*, *2*(3), 47–53.

Bertolino, A., Caforio, G., Blasi, G., De Candia, M., Latorre, V., Petruzzella, V., … Nardini, M. (2004). Interaction of COMT Val108/158 met genotype and olanzapine treatment on prefrontal cortical function in patients with schizophrenia. *American Journal of Psychiatry*, *161*(10), 1798–1805. https://doi.org/10.1176/appi.ajp.161.10.1798

Blanco, C., Hasin, D. S., Wall, M. M., Flórez-Salamanca, L., Hoertel, N., Wang, S.,

… Olfson, M. (2016). Cannabis Use and Risk of Psychiatric Disorders. *JAMA Psychiatry*, *73*(4), 388–395. https://doi.org/10.1001/jamapsychiatry.2015.3229

Bobadilla, L., Vaske, J., & Asberg, K. (2013). Dopamine receptor (D4) polymorphism is related to comorbidity between marijuana abuse and depression. *Addictive Behaviors*, *38*(10), 2555–2562. https://doi.org/10.1016/j.addbeh.2013.05.014

Bouchard, T. J., Lykken, D. T., McGue, M., Segal, N. L., & Tellegen, A. (1990). Sources of human psychological differences: the Minnesota Study of Twins Reared Apart. *Science*, *250*(4978), 223–228. https://doi.org/10.1126/science.2218526

Bovasso, G. B. (2001). Cannabis abuse as a risk factor for depressive symptoms. *American Journal of Psychiatry*, *158*(12), 2033–2037. https://doi.org/10.1176/appi.ajp.158.12.2033

Boyd, J. H., Burke, J. D., Gruenberg, E., Holzer, C. E., Rae, D. S., George, L. K., … Nestadt, G. (1984). Exclusion criteria of DSM-III. A study of co-occurrence of hierarchy-free syndromes. *Archives of General Psychiatry*, *41*(10), 983–989. https://doi.org/10.1001/archpsyc.1984.01790210065008

Brook, D. W., Brook, J. S., Zhang, C., Cohen, P., & Whiteman, M. (2002). Drug use and the risk of major depressive disorder, alcohol dependence, and substance use disorders. *Archives of General Psychiatry*, *59*(11), 1039–1044. https://doi.org/10.1001/archpsyc.59.11.1039

Brook, J. S., Rosen, Z., & Brook, D. W. (2001). The effect of early marijuana use on later anxiety and depressive symptoms. *NYS Psychologist*, *13*, 35–40.

Brook, J. S., Zhang, C., & Brook, D. W. (2011). Developmental trajectories of marijuana use from adolescence to adulthood: personal predictors. *Archives of Pediatrics & Adolescent Medicine*, *165*(1), 55–60. https://doi.org/10.1001/archpediatrics.2010.248.Developmental

Brown, E. S., & Suppes, T. (1998). Mood Symptoms during Corticosteroid Therapy:

A Review. *Harvard Review of Psychiatry*, *5*(5), 239–246. https://doi.org/10.3109/10673229809000307

Bucholz, K. K., Cadoret, R., Cloninger, C. R., Dinwiddie, S. H., Hesselbrock, V. M., Nurnberger, J. I., … Schuckit, M. A. (1994). A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. *Journal of Studies on Alcohol*, *55*(2), 149–158. https://doi.org/http://dx.doi.org/10.15288/jsa.1994.55.149

Buckner, J. D., Crosby, R. D., Silgado, J., Wonderlich, S. A., & Schmidt, N. B. (2012). Immediate antecedents of marijuana use: An analysis from ecological momentary assessment. *Journal of Behavior Therapy and Experimental Psychiatry*, *43*(1), 647–655. https://doi.org/10.1016/j.jbtep.2011.09.010

Buckner, J. D., Crosby, R. D., Wonderlich, S. A., & Schmidt, N. B. (2012). Social anxiety and cannabis use: An analysis from ecological momentary assessment. *Journal of Anxiety Disorders*, *26*(2), 297–304. https://doi.org/10.1016/j.janxdis.2011.12.006

Buckner, J. D., Zvolensky, M. J., Crosby, R. D., Wonderlich, S. A., Ecker, A. H., & Richter, A. (2015). Antecedents and consequences of cannabis use among racially diverse cannabis users: An analysis from Ecological Momentary Assessment. *Drug and Alcohol Dependence*, *147*, 20–25. https://doi.org/10.1016/j.drugalcdep.2014.12.022

Bulik-Sullivan, B. K., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Consortium, R., … Neale, B. M. (2015). An Atlas of Genetic Correlations across Human Diseases and Traits. *bioRxiv*. https://doi.org/10.1101/014498

Bulik-Sullivan, B. K., Loh, P.-R., Finucane, H. K., Ripke, S., Yang, J., Patterson, N., … Consortium, S. W. G. of the P. G. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, *47*(3), 291–295. https://doi.org/doi:10.1038/ng.3211

Burgess, S., & Thompson, S. G. (2011). Avoiding bias from weak instruments in mendelian randomization studies. *International Journal of Epidemiology*, *40*(3), 755–764. https://doi.org/10.1093/ije/dyr036

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. *Ecological Modelling* (2nd ed., Vol. 172). New York: Springer-Verlag. https://doi.org/10.1007/b97636

Carey, C. E., Agrawal, A., Bucholz, K. K., Hartz, S. M., Lynskey, M. T., Nelson, E. C., … Bogdan, R. (2016). Associations between polygenic risk for psychiatric disorders and substance involvement. *Frontiers in Genetics*, *7*. https://doi.org/10.3389/fgene.2016.00149

Caron, C., & Rutter, M. (1991). Comorbidity in child psychopathology: concepts, issues and research strategies. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *32*(7), 1063–1080. https://doi.org/10.1111/j.1469-7610.1991.tb00350.x

Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., … Poulton, R. (2003). Influence of Life Stress on Depression: Moderation by a Polymorphism in the 5-HTT Gene. *Science*, *301*(5631), 386–389. https://doi.org/10.1126/science.1083968

Cerdá, M., Bordelois, P. M., Keyes, K. M., Galea, S., Koenen, K. C., & Pardini, D. (2013). Cumulative and recent psychiatric symptoms as predictors of substance use onset: Does timing matter? *Addiction*, *108*(12), 2119–2128. https://doi.org/10.1111/add.12323

Cerdá, M., Moffitt, T. E., Meier, M. H., Harrington, H. L., Houts, R., Ramrakha, S., … Caspi, A. (2016). Persistent Cannabis Dependence and Alcohol Dependence Represent Risks for Midlife Economic and Social Problems: A Longitudinal Cohort Study. *Clinical Psychological Science*, *4*(6), 1028–1046. https://doi.org/10.1177/2167702616630958

Cerdá, M., Wall, M., Feng, T., Keyes, K. M., Sarvet, A., Schulenberg, J., … Hasin, D. S. (2017). Association of state recreational marijuana laws with adolescent marijuana use. *JAMA Pediatrics*, *171*(2), 142–149. https://doi.org/10.1001/jamapediatrics.2016.3624

Chen, C.-Y., Wagner, F. A., & Anthony, J. C. (2002). Marijuana use and the risk of Major Depressive Episode. Epidemiological evidence from the United States National Comorbidity Survey. *Social Psychiatry and Psychiatric Epidemiology*, *37*(5), 199–206. https://doi.org/10.1007/s00127-002-0541-z

Chen, E., & Miller, G. E. (2013). Socioeconomic Status and Health: Mediating and Moderating Factors. *Annual Review of Clinical Psychology*, *9*(1), 723–749. https://doi.org/10.1146/annurev-clinpsy-050212-185634

Chisholm, D., Sweeny, K., Sheehan, P., Rasmussen, B., Smit, F., Cuijpers, P., & Saxena, S. (2016). Scaling-up treatment of depression and anxiety: A global return on investment analysis. *The Lancet Psychiatry*, *3*(5), 415–424. https://doi.org/10.1016/S2215-0366(16)30024-4

Christensen, R., Kristensen, P. K., Bartels, E. M., Bliddal, H., & Astrup, A. (2007). Efficacy and safety of the weight-loss drug rimonabant: a meta-analysis of randomised trials. *Lancet*, *370*(9600), 1706–1713. https://doi.org/10.1016/S0140-6736(07)61721-8

Clifford, C. A., Hopper, J. L., Fulker, D. W., Murray, R. M., & Rao, D. C. (1984). A genetic and environmental analysis of a twin family study of alcohol use, anxiety and depression. *Genetic Epidemiology*, *1*(1), 63–79. https://doi.org/10.1002/gepi.1370010109

Cohen, J., Cohen, P., West, S. G., & Aiken, L. (2003). *Applied Multiple Regression / Correlation Analysis for the Behavioral Sciences* (3rd ed.). Manway, NJ: Lawrence Erlbaum Associates Publishers. https://doi.org/10.2307/2064799

Compton, W. M., Thomas, Y. F., Stinson, F. S., & Grant, B. F. (2007). Prevalence,

correlates, disability, and comorbidity of DSM-IV drug abuse and dependence in the United States: results from the national epidemiologic survey on alcohol and related conditions. *Archives of General Psychiatry*, *64*(5), 566–576. https://doi.org/10.1001/archpsyc.64.5.566

Copeland, J., Pokorski, I., & Gibson, L. (2017). Overview of Current State-of-the-Art Treatments for Cannabis Use Disorders, and Future Directions. *Current Addiction Reports*, *4*(2), 82–89. https://doi.org/10.1007/s40429-017-0151-1

Coughlin, S. S. (1990). Recall bias in epidemiologic studies. *Journal of Clinical Epidemiology*, *43*(1), 87–91. https://doi.org/10.1016/0895-4356(90)90060-3

Cougle, J. R., Hakes, J. K., Macatee, R. J., Chavarria, J., & Zvolensky, M. J. (2015). Quality of life and risk of psychiatric disorders among regular users of alcohol, nicotine, and cannabis: An analysis of the National Epidemiological Survey on Alcohol and Related Conditions (NESARC). *Journal of Psychiatric Research*, *66–67*, 135–141. https://doi.org/10.1016/j.jpsychires.2015.05.004

Cross-Disorder Group of the Psychiatric Genomics Consortium. (2013). Identification of risk loci with shared effects on five major psychiatric disorders: A genome-wide analysis. *The Lancet*, *381*(9875), 1371–1379. https://doi.org/10.1016/S0140-6736(12)62129-1

Danielsson, A.-K., Lundin, A., Agardh, E., Allebeck, P., & Forsell, Y. (2016). Cannabis use, depression and anxiety: A 3-year prospective population-based study. *Journal of Affective Disorders*, *193*, 103–108. https://doi.org/10.1016/j.jad.2015.12.045

Davey Smith, G., & Hemani, G. (2014). Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Human Molecular Genetics*, *23*(R1), R89-98. https://doi.org/10.1093/hmg/ddu328

De Graaf, R., Radovanovic, M., Van Laar, M., Fairman, B., Degenhardt, L., Aguilar-Gaxiola, S., … Anthony, J. C. (2010). Early cannabis use and estimated risk of

later onset of depression spells: Epidemiologic evidence from the population-based world health organization world mental health survey initiative. *American Journal of Epidemiology*, *172*(2), 149–159. https://doi.org/10.1093/aje/kwq096

Deckman, T., DeWall, C. N., Way, B., Gilman, R., & Richman, S. (2014). Can Marijuana Reduce Social Pain? *Social Psychological and Personality Science*, *5*(2), 131–139. https://doi.org/10.1177/1948550613488949

Degenhardt, L., Coffey, C., Romaniuk, H., Swift, W., Carlin, J. B., Hall, W. D., & Patton, G. C. (2013). The persistence of the association between adolescent cannabis use and common mental disorders into young adulthood. *Addiction*, *108*, 124–133. https://doi.org/10.1111/j.1360-0443.2012.04015.x

Degenhardt, L., Glantz, M., Bharat, C., Peacock, A., Lago, L., Sampson, N., & Kessler, R. C. (2018). The impact of cohort substance use upon likelihood of transitioning through stages of alcohol and cannabis use and use disorder: Findings from the Australian National Survey on Mental Health and Wellbeing. *Drug and Alcohol Review*. https://doi.org/10.1111/dar.12679

Degenhardt, L., Hall, W. D., & Lynskey, M. T. (2001). The relationship between cannabis use, depression and anxiety among Australian adults: Findings from the National Survey of Mental Health and Well-Being. *Social Psychiatry and Psychiatric Epidemiology*, *36*(5), 219–227. https://doi.org/10.1007/s001270170052

Degenhardt, L., Hall, W. D., Lynskey, M. T., Cofey, C., & Patton, G. C. (2012). The association between cannabis use and depression: a review of the evidence. In D. Castle & R. M. Murray (Eds.), *Marijuana and madness: Psychiatry and neurobiology* (2nd ed., pp. 115–128). Cambridge: Cambridge University Press. https://doi.org/http://dx.doi.org/10.1017/CBO9780511706080

Degenhardt, L., Whiteford, H., & Hall, W. D. (2014). The Global Burden of Disease projects: What have we learned about illicit drug use and dependence and their contribution to the global burden of disease? *Drug and Alcohol Review*, *33*(1),

4–12. https://doi.org/10.1111/dar.12088

Demontis, D., Rajagopal, V. M., Als, T. D., Grove, J., Pallesen, J., Hjorthoj, C., … Baekvad-Hansen, M. (2018). Genome-wide association study implicates CHRNA2 in cannabis use disorder. *bioRxiv*. https://doi.org/https://doi.org/10.1101/237321

Di Forti, M., Marconi, A., Carra, E., Fraietta, S., Trotta, A., Bonomo, M., … Murray, R. M. (2015). Proportion of patients in south London with first-episode psychosis attributable to use of high potency cannabis: A case-control study. *The Lancet Psychiatry*, *2*(3), 233–238. https://doi.org/10.1016/S2215-0366(14)00117-5

Edwards, A. C., & Kendler, K. S. (2012). A twin study of depression and nicotine dependence: Shared liability or causal relationship? *Journal of Affective Disorders*, *142*(1–3), 90–97. https://doi.org/10.1016/j.jad.2012.03.048

Egan, M. F., Goldberg, T. E., Kolachana, B. S., Callicott, J. H., Mazzanti, C. M., Straub, R. E., … Weinberger, D. R. (2001). Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(12), 6917–6922. https://doi.org/10.1073/pnas.111134598

Eley, T. C., Sugden, K., Corsico, A., Gregory, A. M., Sham, P., McGuffin, P., … Craig, I. W. (2004). Gene-environment interaction analysis of serotonin system markers with adolescent depression. *Mol Psychiatry*, *9*(10), 908–915. Retrieved from http://dx.doi.org/10.1038/sj.mp.4001546

EMCDDA. (2017). *European Drug Report 2017: Trends and Developments*. Luxembourg: Publications Office of the European Union. https://doi.org/10.2810/88175

Evans, D. M., Gillespie, N. A., & Martin, N. G. (2002). Biometrical genetics. *Biological Psychology*, *61*(1–2), 33–51. https://doi.org/10.1016/S0301-0511(02)00051-0

Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics* (4th

ed.). Harlow: Longman Group Ltd.

Feingold, D., Weiser, M., Rehm, J., & Lev-Ran, S. (2015). The association between cannabis use and mood disorders: A longitudinal study. *Journal of Affective Disorders*, *172*, 211–218. https://doi.org/10.1016/j.jad.2014.10.006

Feinstein, A. R. (1970). The pre-therapeutic classification of co-morbidity in chronic disease. *Journal of Chronic Diseases*, *23*(7), 455–468. https://doi.org/10.1016/0021-9681(70)90054-8

Fergusson, D. M., & Horwood, L. J. (1997). Early onset cannabis use and psychosocial adjustment in young adults. *Addiction*, *92*(3), 279–296. https://doi.org/10.1111/j.1360-0443.1997.tb03198.x

Fergusson, D. M., Lynskey, M. T., & Horwood, L. J. (1996). Childhood Sexual Abuse and Psychiatric Disorder in Young Adulthood: I. Prevalence of Sexual Abuse and Factors Associated with Sexual Abuse. *Journal of the American Academy of Child & Adolescent Psychiatry*, *35*(10), 1355–1364. https://doi.org/10.1097/00004583-199610000-00023

Ferrari, A. J., Charlson, F. J., Norman, R. E., Patten, S. B., Freedman, G., Murray, C. J. L., … Whiteford, H. a. (2013). Burden of Depressive Disorders by Country, Sex, Age, and Year: Findings from the Global Burden of Disease Study 2010. *PLoS Medicine*, *10*(11). https://doi.org/10.1371/journal.pmed.1001547

Field, A., Miles, J., & Field, Z. (2012). Logistic Regression. In *Discovaring statistics using R* (pp. 312–358). London: SAGE Publications. https://doi.org/10.1111/insr.12011_21

Fisher, R. A. (1918). XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh*, *52*(2), 399–433. https://doi.org/10.1017/S0080456800012163

Freeman, T. P., & Winstock, A. R. (2015). Examining the profile of high-potency cannabis and its association with severity of cannabis dependence.

Psychological Medicine, *45*(15), 3181–3189. https://doi.org/10.1017/S0033291715001178

Fu, Q., Heath, A. C., Bucholz, K. K., Nelson, E., Goldberg, J., Lyons, M. J., … Eisen, S. A. (2002). Shared genetic risk of major depression, alcohol dependence, and marijuana dependence: contribution of antisocial personality disorder in men. *Archives of General Psychiatry*, *59*(12), 1125–1132. https://doi.org/10.1001/archpsyc.59.12.1125

Furlong, M., McGilloway, S., Bywater, T., Hutchings, J., Smith, S. M., & Donnelly, M. (2013). Cochrane Review: Behavioural and cognitive-behavioural group-based parenting programmes for early-onset conduct problems in children aged 3 to 12 years (Review). *Evidence-Based Child Health: A Cochrane Review Journal*, *8*(2), 318–692. https://doi.org/10.1002/ebch.1905

Gage, S. H., Hickman, M., Heron, J., Munafò, M. R., Lewis, G., Macleod, J., & Zammit, S. (2015). Associations of cannabis and cigarette use with psychotic experiences at age 18: Findings from the Avon Longitudinal Study of Parents and Children. *Psychological Medicine*, *44*(16), 3435–3444. https://doi.org/10.1017/S0033291714000531

Galton, F. (1875). *English men of science: Their nature and nurture*. New York: D. Appleton and Co.

Gates, P. J., Sabioni, P., Copeland, J., Le Foll, B., & Gowing, L. (2016). Psychosocial interventions for cannabis use disorder. *Cochrane Database of Systematic Reviews*. https://doi.org/10.1002/14651858.CD005336.pub4

Georgiades, K., & Boyle, M. H. (2007). Adolescent tobacco and cannabis use: young adult outcomes from the Ontario Child Health Study. *Journal of Child Psychology and Psychiatry*, *48*(7), 724–731.

Gilman, S. E., Kawachi, I., Fitzmaurice, G. M., & Buka, S. L. (2002). Socioeconomic status in childhood and the lifetime risk of major depression. *Int J Epidemiol*,

31(2), 359–367. https://doi.org/10.1093/ije/31.2.359

Goldenberg, M., Reid, M. W., IsHak, W. W., & Danovitch, I. (2017). The impact of cannabis and cannabinoids for medical conditions on health-related quality of life: A systematic review and meta-analysis. *Drug and Alcohol Dependence*. https://doi.org/10.1016/j.drugalcdep.2016.12.030

Gorzalka, B. B., Hill, M. N., & Hillard, C. J. (2008). Regulation of endocannabinoid signaling by stress: Implications for stress-related affective disorders. *Neuroscience and Biobehavioral Reviews*, *32*(6), 1152–1160. https://doi.org/10.1016/j.neubiorev.2008.03.004

Greenberg, P. E., Fournier, A.-A., Sisitsky, T., Pike, C. T., & Kessler, R. C. (2015). The Economic Burden of Adults with Major Depressive Disorder in the United States (2005 and 2010). *Journal of Clinical Psychiatry*, *76*(2), 155–162. https://doi.org/10.4088/JCP.14m09298

Hakkarainen, P., Frank, V. A., Barratt, M. J., Dahl, H. V., Decorte, T., Karjalainen, K., … Werse, B. (2015). Growing medicine: Small-scale cannabis cultivation for medical purposes in six different countries. *International Journal of Drug Policy*, *26*(3), 250–256. https://doi.org/10.1016/j.drugpo.2014.07.005

Hall, W. D. (2015). What has research over the past two decades revealed about the adverse health effects of recreational cannabis use? *Addiction*, *110*(1), 19–35. https://doi.org/10.1111/add.12703

Hall, W. D., & Degenhardt, L. (2009). Adverse health effects of non-medical cannabis use. *The Lancet*, *374*(9698), 1383–1391. https://doi.org/10.1016/S0140-6736(09)61037-0

Hamilton, J. P., Siemer, M., & Gotlib, I. H. (2008). Amygdala volume in major depressive disorder: a meta-analysis of magnetic resonance imaging studies. *Molecular Psychiatry*, *13*(11), 993–1000. https://doi.org/10.1038/mp.2008.57

Handsaker, R. E., Van Doren, V., Berman, J. R., Genovese, G., Kashin, S., Boettger,

L. M., & McCarroll, S. a. (2015). Large multiallelic copy number variations in humans. *Nature Genetics*, *47*, 1–10. https://doi.org/10.1038/ng.3200

Hankin, B. L., Fraley, R. C., Lahey, B. B., & Waldman, I. D. (2005). Is depression best viewed as a continuum or discrete category? A taxometric analysis of childhood and adolescent depression in a population-based sample. *Journal of Abnormal Psychology*, *114*(1), 96–110. https://doi.org/10.1037/0021-843X.114.1.96

Harder, V. S., Morral, A. R., & Arkes, J. (2006). Marijuana use and depression among adults: Testing for causal associations. *Addiction*, *101*(10), 1463–1472. https://doi.org/10.1111/j.1360-0443.2006.01545.x

Harder, V. S., Stuart, E. A., & Anthony, J. C. (2008). Adolescent cannabis problems and young adult depression: Male-female stratified propensity score analyses. *American Journal of Epidemiology*, *168*, 592–601. https://doi.org/10.1093/aje/kwn184

Hasin, D. S., O'Brien, C. P., Auriacombe, M., Borges, G., Bucholz, K., Budney, A., … Grant, B. F. (2013). DSM-5 criteria for substance use disorders: Recommendations and rationale. *American Journal of Psychiatry*, *170*(8), 834–851. https://doi.org/10.1176/appi.ajp.2013.12060782

Hasin, D. S., Saha, T. D., Kerridge, B. T., Goldstein, R. B., Chou, S. P., Zhang, H. T., … Grant, B. F. (2015). Prevalence of Marijuana Use Disorders in the United States Between 2001-2002 and 2012-2013. *JAMA Psychiatry*, *72*(12), 1235–1242. https://doi.org/10.1001/jamapsychiatry.2015.1858

Haug, S., López Núñez, C., Becker, J., Gmel, G., & Schaub, M. P. (2014). Predictors of the late onset of cannabis and other drug use in male young adults. *BMC Public Health*, *14*(1), 1202–1217. https://doi.org/10.1186/1471-2458-14-1202

Heath, A. C., Kessler, R. C., Neale, M. C., Hewitt, J. K., Eaves, L. J., & Kendler, K. S. (1993). Testing hypotheses about direction of causation using cross-sectional

family data. *Behavior Genetics*, *23*(1), 29–50. https://doi.org/10.1007/BF01067552

Heath, A. C., Martin, N. G., Lynskey, M. T., Todorov, A. A., & Madden, P. A. F. (2002). Estimating two-stage models for genetic influences on alcohol, tobacco or drug use initiation and dependence vulnerability in twin and family data. *Twin Research*, *5*(2), 113–124. https://doi.org/10.1375/twin.5.2.113

Heath, A. C., Todorov, A. a, Nelson, E. C., Madden, P. a F., Bucholz, K. K., & Martin, N. G. (2002). Gene-environment interaction effects on behavioral variation and risk of complex disorders: the example of alcoholism and other psychiatric disorders. *Twin Research : The Official Journal of the International Society for Twin Studies*, *5*(1), 30–37. https://doi.org/10.1375/twin.5.1.30

Heim, C., & Nemeroff, C. B. (2001). The role of childhood trauma in the neurobiology of mood and anxiety disorders: Preclinical and clinical studies. *Biological Psychiatry*, *49*(12), 1023–1039. https://doi.org/10.1016/S0006-3223(01)01157-X

Herbert, J., Goodyer, I. M., Grossman, A. B., Hastings, M. H., de Kloet, E. R., Lightman, S. L., … Seckl, J. R. (2006). Do corticosteroids damage the brain? *Journal of Neuroendocrinology*, *18*(6), 393–411. https://doi.org/10.1111/j.1365-2826.2006.01429.x

Herman, J. P., Ostrander, M. M., Mueller, N. K., & Figueiredo, H. (2005). Limbic system mechanisms of stress regulation: Hypothalamo-pituitary- adrenocortical axis. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *29*(8), 1201–1213. https://doi.org/10.1016/j.pnpbp.2005.08.006

Hesselbrock, M., Easton, C., Bucholz, K. K., Schuckit, M., Hesselbrock, V., M., H., … V., H. (1999). A validity study of the SSAGA - A comparison with the SCAN. *Addiction*. https://doi.org/https://doi.org/10.1046/j.1360-0443.1999.94913618.x

Hill, M. N., Carrier, E. J., McLaughlin, R. J., Morrish, A. C., Meier, S. E., Hillard, C. J., & Gorzalka, B. B. (2008). Regional alterations in the endocannabinoid system in

an animal model of depression: Effects of concurrent antidepressant treatment. *Journal of Neurochemistry*, *106*(6), 2322–2336. https://doi.org/10.1111/j.1471-4159.2008.05567.x

Hill, M. N., Miller, G. E., Carrier, E. J., Gorzalka, B. B., & Hillard, C. J. (2009). Circulating endocannabinoids and N-acyl ethanolamines are differentially regulated in major depression and following exposure to social stress. *Psychoneuroendocrinology*, *34*(8), 1257–1262. https://doi.org/10.1016/j.psyneuen.2009.03.013

Hines, L. A., Morley, K. I., Rijsdijk, F., Strang, J., Agrawal, A., Nelson, E. C., … Lynskey, M. T. (2018). Overlap of heritable influences between cannabis use disorder, frequency of use and opportunity to use cannabis: Trivariate twin modelling and implications for genetic design. *Psychological Medicine*, *48*, 2486–2793. https://doi.org/10.1017/S0033291718000478

Hirvonen, J., Goodwin, R. S., Li, C. T., Terry, G. E., Zoghbi, S. S., Morse, C., … Innis, R. B. (2012). Reversible and regionally selective downregulation of brain cannabinoid CB1receptors in chronic daily cannabis smokers. *Molecular Psychiatry*, *17*(6), 642–649. https://doi.org/10.1038/mp.2011.82

Hoch, E., Bühringer, G., Pixa, A., Dittmer, K., Henker, J., Seifert, A., & Wittchen, H. U. (2014). CANDIS treatment program for cannabis use disorders: Findings from a randomized multi-site translational trial. *Drug and Alcohol Dependence*, *134*, 185–193. https://doi.org/10.1016/j.drugalcdep.2013.09.028

Hodgson, K., Almasy, L., Knowles, E. E. M., Kent, J. W., Curran, J. E., Dyer, T. D., … Glahn, D. C. (2016). The genetic basis of the comorbidity between cannabis use and major depression. *Addiction*, *12*(2), 113–123. https://doi.org/10.1111/add.13558

Holsboer, F. (2000). The corticosteroid receptor hypothesis of depression. *Neuropsychopharmacology*, *23*(5), 477–501. https://doi.org/10.1016/S0893-133X(00)00159-7

Hopper, J. L. (2002). The Australian Twin Registry. *Twin Research : The Official Journal of the International Society for Twin Studies*, *5*(5), 329–336. https://doi.org/10.1375/twin.5.5.329

Hopper, J. L., Foley, D. L., White, P. A., & Pollaers, V. (2013). Australian Twin Registry: 30 Years of Progress. *Twin Research and Human Genetics*, *16*(1), 34–42. https://doi.org/10.1017/thg.2012.121

Hopper, J. L., Treloar, S. a, de Klerk, N. H., & Morley, R. (2006). Australian Twin Registry: a nationally funded resource for medical and scientific research, incorporating match and WATCH. *Twin Research and Human Genetics : The Official Journal of the International Society for Twin Studies*, *9*(6), 707–711. https://doi.org/10.1375/183242706779462507

Hornor, G. (2018). Child Sexual Abuse: Consequences and Implications. *Journal of Pediatric Health Care*, *24*(6), 358–364. https://doi.org/10.1016/j.pedhc.2009.07.003

Horwitz, A. V, Videon, T. M., Schmitz, M. F., & Davis, D. (2003). Rethinking twins and environments: possible social sources for assumed genetic influences in twin research. *Journal of Health and Social Behaviour*, *44*(2), 111–129. https://doi.org/10.2307/1519802

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Johnson, T. P., & Mott, J. a. (2001). The reliability of self-reported age of onset of tobacco, alcohol and illicit drug use. *Addiction*, *96*(8), 1187–1198. https://doi.org/10.1080/09652140120060770

Kandel, D. B., Yamaguchi, K., & Chen, K. (1992). Stages of progression in drug involvement from adolescence to adulthood: further evidence for the gateway theory. *Journal of Studies on Alcohol*, *53*(5), 447–457. https://doi.org/https://doi.org/10.15288/jsa.1992.53.447

Kendler, K. S. (1983). Overview: a current perspective on twin studies of schizophrenia. *American Journal of Psychiatry*, *140*(11), 1413–1425. https://doi.org/10.1176/ajp.140.11.1413

Kendler, K. S., Aggen, S. H., Tambs, K., & Reichborn-Kjennerud, T. E. D. (2006). Illicit psychoactive substance use, abuse and dependence in a population-based sample of Norwegian twins. *Psychological Medicine*, *36*(7), 955–962. https://doi.org/10.1017/S0033291706007720

Kendler, K. S., Gatz, M., Gardner, C. O., & Pedersen, N. L. (2006). A Swedish national twin study of lifetime major depression. *American Journal of Psychiatry*, *163*(1), 109–114. https://doi.org/10.1176/appi.ajp.163.1.109

Kendler, K. S., Kessler, R. C., Walters, E. E., MacLean, C., Neale, M. C., Heath, A. C., & Eaves, L. J. (1995). Stressful life events, genetic liability, and onset of an episode of major depression in women. *American Journal of Psychiatry*, *152*(6), 833–842. https://doi.org/10.1176/ajp.152.6.833

Kendler, K. S., Kuhn, J. W., Vittum, J., Prescott, C. A., & Riley, B. (2005). The interaction of stressful life events and a serotonin transporter polymorphism in the prediction of episodes of major depression: A replication. *Archives of General Psychiatry*, *62*(5), 529–535. https://doi.org/http://dx.doi.org/10.1001/archpsyc.62.5.529

Kendler, K. S., Martin, N. G., Heath, A. C., & Eaves, L. J. (1995). Self-report psychiatric symptoms in twins and their nontwin relatives: Are twins different? *American Journal of Medical Genetics - Neuropsychiatric Genetics*, *60*(6), 588–591. https://doi.org/10.1002/ajmg.1320600622

Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C., & Eaves, L. J. (1993). A test of the equal-environment assumption in twin studies of psychiatric illness. *Behavior Genetics*, *23*(1), 21–27. https://doi.org/10.1007/BF01067551

Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C., & Eaves, L. J. (1994).

Parental treatment and the equal environment assumption in twin studies of psychiatric illness. *Psychological Medicine*, *24*(3), 579–590. https://doi.org/https://doi.org/10.1017/S0033291700027732

Kendler, K. S., Neale, M. C., MacLean, C. J., Heath, A. C., Eaves, L. J., & Kessler, R. C. (1993). Smoking and major depression. A causal analysis. *Archives of General Psychiatry*, *50*(1), 36–43. https://doi.org/10.1001/archpsyc.1993.01820130038007

Kendler, K. S., Pedersen, N. L., Farahmand, B. Y., & Persson, P. G. (1996). The treated incidence of psychotic and affective illness in twins compared with population expectation: a study in the Swedish Twin and Psychiatric Registries. *Psychological Medicine*, *26*(6), 1135–1144. https://doi.org/10.1017/S0033291700035856

Kessler, R. C. (2004). The epidemiology of dual diagnosis. *Biological Psychiatry*, *56*(10), 730–737. https://doi.org/10.1016/j.biopsych.2004.06.034

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. *Archives of General Psychiatry*, *62*(6), 593–602. https://doi.org/10.1001/archpsyc.62.6.593

Kessler, R. C., & Bromet, E. J. (2013). The epidemiology of depression across cultures. *Annual Review of Public Health*, *34*, 119–138. https://doi.org/10.1146/annurev-publhealth-031912-114409

Kessler, R. C., McGonagle, K. a, Zhao, S., Nelson, C. B., Hughes, M., Eshleman, S., … Kendler, K. S. (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. *Archives of General Psychiatry*, *51*(1), 8–19. https://doi.org/10.1001/archpsyc.1994.03950010008002

Keyes, K. M., Schulenberg, J. E., O'Malley, P. M., Johnston, L. D., Bachman, J. G.,

Li, G., & Hasin, D. S. (2011). The social norms of birth cohorts and adolescent marijuana use in the United States, 1976-2007. *Addiction*, *106*(10), 1790–1800. https://doi.org/10.1111/j.1360-0443.2011.03485.x

Klein, D. N., & Riso, L. P. (1993). Psychiatric disorders: problems of boundaries and comorbidity. In C. G. Costello (Ed.), *Basic issues in psychopathology* (pp. 19–66). New York: Guildford.

Kukolja, J., Schlapfer, T. E., Keysers, C., Klingmuller, D., Maier, W., Fink, G. R., & Hurlemann, R. (2008). Modeling a Negative Response Bias in the Human Amygdala by Noradrenergic-Glucocorticoid Interactions. *Journal of Neuroscience*, *28*(48), 12868–12876. https://doi.org/10.1523/JNEUROSCI.3592-08.2008

Lawlor, D. A., Harbord, R. M., Sterne, J. A. C., Timpson, N., & Davey Smith, G. (2008). Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Statistics in Medicine*, *27*(8), 1133–1163. https://doi.org/10.1002/sim.3034

Lev-Ran, S., & Feingold, D. (2017). Cannabis use and its association to mental illness: a focus on mood and anxiety disorders. In V. R. Preedy (Ed.), *Handbook of Cannabis and Related Pathologies* (1st ed., pp. 298–307). Academic Press. https://doi.org/10.1016/B978-0-12-800756-3.00035-1

Lev-Ran, S., Roerecke, M., Le Foll, B., George, T. P., McKenzie, K., & Rehm, J. (2014). The association between cannabis use and depression: a systematic review and meta-analysis of longitudinal studies. *Psychological Medicine*, *44*(4), 797–810. https://doi.org/10.1017/S0033291713001438

Levine, A., Huang, Y. Y., Drisaldi, B., Griffin, E. A., Pollak, D. D., Xu, S., … Kandel, E. R. (2011). Molecular mechanism for a gateway drug: Epigenetic changes initiated by nicotine prime gene expression by cocaine. *Science Translational Medicine*, *3*(107), 1–7. https://doi.org/10.1126/scitranslmed.3003062

Lin, N., Eisen, S. A., Scherrer, J. F., Goldberg, J., True, W. R., Lyons, M. J., & Tsuang, M. T. (1996). The influence of familial and non-familial factors on the association between major depression and substance abuse/dependence in 1874 monozygotic male twin pairs. *Drug and Alcohol Dependence*, *43*(1–2), 49–55. https://doi.org/10.1016/S0376-8716(96)01287-2

Loehlin, J. C., & Nichols, R. C. (1976). *Heredity, Environment, & Personality: A Study of 850 Sets of Twins.* Austin, Texas: University of Texas Press.

Lynskey, M. T., Agrawal, A., Henders, A., Nelson, E. C., Madden, P. A. F., & Martin, N. G. (2012). An Australian Twin Study of Cannabis and Other Illicit Drug Use and Misuse, and Other Psychopathology. *Twin Research and Human Genetics*, *15*(5), 631–641. https://doi.org/10.1017/thg.2012.41

Lynskey, M. T., Glowinski, A. L., Todorov, A. A., Bucholz, K. K., Madden, P. A. F., Nelson, E. C., … Heath, A. C. (2004). Major depressive disorder, suicidal ideation, and suicide attempt in twins discordant for cannabis dependence and early-onset cannabis use. *Archives of General Psychiatry*, *61*(10), 1026–1032. https://doi.org/10.1016/S0084-3970(08)70083-5

Lynskey, M. T., & Hall, W. D. (2000). The effects of adolescent cannabis use on educational attainment: a review. *Addiction*, *95*(11), 1621–1630. https://doi.org/10.1046/j.1360-0443.2000.951116213.x

Lynskey, M. T., Heath, A. C., Nelson, E. C., Bucholz, K. K., Madden, P. A. F., Slutske, W. S., … Martin, N. G. (2002). Genetic and environmental contributions to cannabis dependence in a national young adult twin sample. *Psychological Medicine*, *32*(2), 195–207. https://doi.org/10.1017/S0033291701005062

Lytton, H. (1977). Do parents create, or respond to, differences in twins? *Developmental Psychology*, *13*(5), 456–459. https://doi.org/10.1037/0012-1649.13.5.456

MacCallum, R., Wegener, D., Uchino, B., & Fabrigar, L. (1993). The Problem of

Equivalent Models in Applications of Covariance Structure Analysis. *Psychological Bulletin*, *114*(1), 185–199. https://doi.org/10.1037/0033-2909.114.1.185

MacCoun, R. J., & Reuter, P. (2011). Assessing Drug Prohibition and Its Alternatives: A Guide for Agnostics. *Annual Review of Law and Social Science*, *7*(1), 61–78. https://doi.org/10.1146/annurev-lawsocsci-102510-105442

Mackay, T. F. C. (2014). Epistasis and Quantitative Traits: Using Model Organisms to Study Gene-Gene Interactions. *Nature Reviews. Genetics*, *15*(1), 22–33. https://doi.org/10.1038/nrg3627

Macleod, M. R., Michie, S., Roberts, I., Dirnagl, U., Chalmers, I., Ioannidis, J. P. A., … Glasziou, P. (2014). Biomedical research: increasing value, reducing waste. *The Lancet*, *383*(9912), 101–104. https://doi.org/10.1016/S0140-6736(13)62329-6

Maes, H. H., Neale, M. C., Kendler, K. S., Hewitt, J. K., Silberg, J. L., Foley, D. L., … Eaves, L. J. (1998). Assortative mating for major psychiatric diagnoses in two population-based samples. *Psychological Medicine*, *28*(6), 1389–1401. https://doi.org/10.1017/S0033291798007326

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., … Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*, *461*(7265), 747–753. Retrieved from http://dx.doi.org/10.1038/nature08494

Manrique-Garcia, E., Zammit, S., Dalman, C., Hemmingsson, T., & Allebeck, P. (2012). Cannabis use and depression: a longitudinal study of a national cohort of Swedish conscripts. *BMC Psychiatry*, *12*, 112. https://doi.org/10.1186/1471-244X-12-112

Marmorstein, N. R., & Iacono, W. G. (2011). Explaining associations between cannabis use disorders in adolescence and later major depression: A test of the

psychosocial failure model. *Addictive Behaviors*, *36*(7), 773–776. https://doi.org/10.1016/j.addbeh.2011.02.006

Martins, S. S., & Gorelick, D. A. (2011). Conditional substance abuse and dependence by diagnosis of mood or anxiety disorder or schizophrenia in the U.S. population. *Drug and Alcohol Dependence*, *119*(1–2), 28–36. https://doi.org/10.1016/j.drugalcdep.2011.05.010

Matheny, A. P., Wilson, R. S., & Dolan, A. B. (1976). Relations between twins' similarity of appearance and behavioral similarity: Testing an assumption. *Behavior Genetics*, *6*(3), 343–351. https://doi.org/10.1007/BF01065729

Mather, J. K., & Jinks, J. L. (1982). *Biometrical Genetics: the Study of Continuous Variation* (3rd ed.). London: Chapman Hall.

Mathews, C. A., & Reus, V. I. (2001). Assortative mating in the affective disorders: A systematic review and meta-analysis. *Comprehensive Psychiatry*, *42*(4), 257–262. https://doi.org/10.1053/comp.2001.24575

Mathews, R. R. S., Hall, W. D., & Gartner, C. E. (2011). Depression and psychological distress in tobacco smokers and people with cannabis dependence in the National Survey of Mental Health and wellbeing. *Medical Journal of Australia*, *195*(3), S12–S15.

McAdams, T. A., Neiderhiser, J. M., Rijsdijk, F. V, Narusyte, J., Lichtenstein, P., & Eley, T. C. (2014). Accounting for genetic and environmental confounds in associations between parent and child characteristics: a systematic review of children-of-twins studies. *Psychological Bulletin*, *140*(4), 1138–1173. https://doi.org/10.1037/a0036416

McGrath, J., Saha, S., Chant, D., & Welham, J. (2008). Schizophrenia: A concise overview of incidence, prevalence, and mortality. *Epidemiologic Reviews*, *30*(1), 67–76. https://doi.org/10.1093/epirev/mxn001

Mcketin, R., Dunlop, A. J., Holland, R. M., Sutherland, R. A., Baker, A. L., Salmon, A.

M., & Hudson, S. L. (2013). Treatment outcomes for methamphetamine users receiving outpatient counselling from the Stimulant Treatment Program in Australia. *Drug and Alcohol Review*, *32*(1), 80–87. https://doi.org/10.1111/j.1465-3362.2012.00471.x

Mechoulam, R., & Parker, L. A. (2013). The Endocannabinoid System and the Brain. *Annual Review of Psychology*, *64*(1), 21–47. https://doi.org/10.1146/annurev-psych-113011-143739

Medland, S., & Hatemi, P. (2009). Political science, biometric theory, and twin studies: A methodological introduction. *Political Analysis*, *17*(2), 191–214. https://doi.org/10.1093/pan/mpn016

Mendelson, T., & Letourneau, E. J. (2015). Parent-Focused Prevention of Child Sexual Abuse. *Prevention Science*, *16*(6), 844–852. https://doi.org/10.1007/s11121-015-0553-z

Merikangas, K. R., Li, J. J., Stipelman, B., Yu, K., Fucito, L., Swendsen, J., & Zhang, H. (2009). The familial aggregation of cannabis use disorders. *Addiction*, *104*(4), 622–629. https://doi.org/10.1111/j.1360-0443.2008.02468.x

Milne, B. J., Caspi, A., Harrington, H., Poulton, R., Rutter, M., & Moffitt, T. E. (2009). Predictive value of family history on severity of illness: The case for depression, anxiety, alcohol dependence, and drug dependence. *Archives of General Psychiatry*, *66*(7), 738–747. https://doi.org/10.1001/archgenpsychiatry.2009.55

Minica, C. C., Dolan, C. V, Boomsma, D. I., de Geus, E., & Neale, M. C. (2017). Extending Causality Tests With Genetic Instruments: An Integration Of Mendelian Randomization And The Classical Twin Design. *bioRxiv*. https://doi.org/10.1101/134585

Mohammadi, D. (2014). Addiction services in England: in need of an intervention. *The Lancet Psychiatry*, *1*(6), 421–422. https://doi.org/10.1016/S2215-0366(14)00015-7

Morris-Yates, A., Andrews, G., Howie, P., & Henderson, S. (1990). Twins: A test of the equal environments assumption. *Acta Psychiatrica Scandinavica*, *81*(4), 322–326. https://doi.org/10.1111/j.1600-0447.1990.tb05457.x

Moss, H. B., Chen, C. M., & Yi, H. ye. (2014). Early adolescent patterns of alcohol, cigarettes, and marijuana polysubstance use and young adult substance use outcomes in a nationally representative sample. *Drug and Alcohol Dependence*, *136*(1), 51–62. https://doi.org/10.1016/j.drugalcdep.2013.12.011

Muñoz, R. F., Cuijpers, P., Smit, F., Barrera, A. Z., & Leykin, Y. (2010). Prevention of Major Depression. *Annual Review of Clinical Psychology*, *6*(1), 181–212. https://doi.org/10.1146/annurev-clinpsy-033109-132040

Murray, L., & Cooper, P. (1997). Effects of postnatal depression on infant development. *Archives of Disease in Childhood*, *77*(2), 99–101. https://doi.org/http://dx.doi.org/10.1136/adc.77.2.99

Nanni, V., Uher, R., & Danese, A. (2012). Childhood maltreatment predicts unfavorable course of illness and treatment outcome in depression: A meta-analysis. *American Journal of Psychiatry*, *169*(2), 141–151. https://doi.org/10.1176/appi.ajp.2011.11020335

Neale, M. C., Eaves, L. J., & Kendler, K. S. (1994). The power of the classical twin study to resolve variation in threshold traits. *Behavior Genetics*, *24*(3), 239–258. https://doi.org/10.1007/BF01067191

Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R. M., … Boker, S. M. (2016). OpenMx 2.0: Extended structural equation and statistical modeling. *Psychometrika*, *81*(2), 535–549. https://doi.org/10.1007/s11336-014-9435-8

Neale, M. C., & Kendler, K. S. (1995). Models of comorbidity for multifactorial disorders. *American Journal of Human Genetics*, *57*(4), 935–953.

Neale, M. C., & Maes, H. H. M. (2004). *Methodology for Genetic Studies of Twins*

and Families. Dordrecht: Kluwer Academic Publishers B.V.

Neale, M. C., Røysamb, E., & Jacobson, K. (2006). Multivariate genetic analysis of sex limitation and G x E interaction. *Twin Research and Human Genetics : The Official Journal of the International Society for Twin Studies*, *9*(4), 481–489. https://doi.org/10.1375/twin.9.4.481

Nelson, E. C., Heath, A. C., Lynskey, M. T., Bucholz, K. K., Madden, P. A. F., Statham, D. J., & Martin, N. G. (2006). Childhood sexual abuse and risks for licit and illicit drug-related outcomes: a twin study. *Psychological Medicine*, *36*(10), 1473–1483. https://doi.org/10.1017/S0033291706008397

Nesvåg, R., Knudsen, G. P., Bakken, I. J., Høye, A., Ystrom, E., Surén, P., … Reichborn-Kjennerud, T. (2015). Substance use disorders in schizophrenia, bipolar disorder, and depressive illness: a registry-based study. *Social Psychiatry and Psychiatric Epidemiology*, *50*(8), 1267–1276. https://doi.org/10.1007/s00127-015-1025-2

Ni, G., Moser, G., Wray, N. R., & Lee, S. H. (2017). Estimation of genetic correlation using linkage disequilibrium score regression and genomic restricted maximum likelihood. *bioRxiv*. https://doi.org/https://doi.org/10.1101/194019

Nolte, I. M., van der Most, P. J., Alizadeh, B. Z., de Bakker, P. I. W., Boezen, H. M., Bruinenberg, M., … Snieder, H. (2017). Missing heritability: is the gap closing? An analysis of 32 complex traits in the Lifelines Cohort Study. *European Journal Of Human Genetics*, *25*(7), 877–885. https://doi.org/doi:10.1038/ejhg.2017.50

Nutt, D. J., King, L. A., & Phillips, L. D. (2010). Drug harms in the UK: A multicriteria decision analysis. *The Lancet*, *376*(9752), 1558–1565. https://doi.org/10.1016/S0140-6736(10)61462-6

Ogden, T., & Hagen, K. A. (2008). Treatment effectiveness of parent management training in Norway: A randomized controlled trial of children with conduct problems. *Journal of Consulting and Clinical Psychology*, *76*(4), 607–621.

https://doi.org/http://dx.doi.org/10.1037/0022-006X.76.4.607

Olesen, J., Gustavsson, A., Svensson, M., Wittchen, H. U., & Jönsson, B. (2012). The economic cost of brain disorders in Europe. *European Journal of Neurology*, *19*(1), 155–162. https://doi.org/10.1111/j.1468-1331.2011.03590.x

Pacek, L. R., Martins, S. S., & Crum, R. M. (2013). The bidirectional relationships between alcohol, cannabis, co-occurring alcohol and cannabis use disorders with major depressive disorder: Results from a national sample. *Journal of Affective Disorders*, *148*(2–3), 188–195. https://doi.org/10.1016/j.jad.2012.11.059

Pacula, R. L., & Smart, R. (2017). Medical Marijuana and Marijuana Legalization. *Annual Review of Clinical Psychology*, *13*(1), 397–419. https://doi.org/10.1146/annurev-clinpsy-032816-045128

Palmer, R. H. C., Brick, L., Nugent, N. R., Bidwell, L. C., Mcgeary, J. E., Knopik, V. S., & Keller, M. C. (2015). Examining the role of common genetic variants on alcohol, tobacco, cannabis and illicit drug dependence: Genetics of vulnerability to drug dependence. *Addiction*, *110*(3), 530–537. https://doi.org/10.1111/add.12815

Palmer, R. H. C., Button, T. M., Rhee, S. H., Corley, R. P., Young, S. E., Stallings, M. C., … Hewitt, J. K. (2012). Genetic etiology of the common liability to drug dependence: Evidence of common and specific mechanisms for DSM-IV dependence symptoms. *Drug and Alcohol Dependence*, *123*(S), S24–S32. https://doi.org/10.1016/j.drugalcdep.2011.12.015

Pam, A., Kemker, S. S., Ross, C. A., & Golden, R. (1996). The “equal environments assumption” in MZ-DZ twin comparisons: An untenable premise of psychiatric genetics? *Acta Geneticae Medicae et Gemellologiae: Twin Reseaech*, *45*(3), 349–360. https://doi.org/https://doi.org/10.1017/S0001566000000945

Pariante, C. M. (2017). Why are depressed patients inflamed? A reflection on 20

years of research on depression, glucocorticoid resistance and inflammation. *European Neuropsychopharmacology*, *27*(6), 554–559. https://doi.org/10.1016/j.euroneuro.2017.04.001

Pariante, C. M., & Lightman, S. L. (2008). The HPA axis in major depression: classical theories and new developments. *Trends in Neurosciences*, *31*(9), 464–468. https://doi.org/10.1016/j.tins.2008.06.006

Parsons, L. H., & Hurd, Y. L. (2015). Endocannabinoid signalling in reward and addiction. *Nature Reviews Neuroscience*, *16*(10), 579–594. https://doi.org/10.1038/nrn4004

Patel, S., Hill, M. N., & Hillard, C. J. (2014). Effects of Phytocannabinoids on Anxiety, Modd, and the Endocrine System. In R. Pertwee (Ed.), *Handbook of Cannabis* (pp. 189–207). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199662685.003.0010

Patel, S., & Hillard, C. J. (2009). Role of endocannabinoid signaling in anxiety and depression. *Current Topics in Behavioral Neurosciences*, *1*, 347–371. https://doi.org/10.1007/978-3-540-88955-7_14

Paton, S., Kessler, R. C., & Kandel, D. (1977). Depressive mood and adolescent illicit drug use: a longitudinal analysis. *Journal of Genetic Psychology*, *131*, 267–289. https://doi.org/10.1080/00221325.1977.10533299

Patten, S. B., Williams, J. V. A., Lavorato, D. H., Wang, J. L., McDonald, K., & Bulloch, A. G. M. (2015). Descriptive epidemiology of major depressive disorder in Canada in 2012. *Canadian Journal of Psychiatry*, *60*(1), 23–30. https://doi.org/10.1177/070674371506000106

Pearson, K. (1900). *The Grammar of Science*. London: A&C Black.

Pearson, K., & Lee, A. (1901). On the inheritance of characters not capable of exact quantitative measurement. *Philosophical Transactions of the Royal Society of London*, *195*, 79–150. https://doi.org/10.1098/rsta.1900.0024

Pedersen, W. (2008). Does cannabis use lead to depression and suicidal behaviours? A population-based longitudinal study. *Acta Psychiatrica Scandinavica*, *118*(5), 395–403. https://doi.org/10.1111/j.1600-0447.2008.01259.x

Pickrell, J. K., Berisa, T., Liu, J. Z., Segurel, L., Tung, J. Y., & Hinds, D. A. (2016). Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet*, *48*(7), 709–717. Retrieved from http://dx.doi.org/10.1038/ng.3570

Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2013). *Behavioral Genetics* (6th ed.). New York: Worth Publishers.

Plomin, R., Haworth, C. M. A., & Davis, O. S. P. (2009). Common disorders are quantitative traits. *Nature Reviews Genetics*, *10*(12), 872–878. https://doi.org/10.1038/nrg2670

Plomin, R., Willerman, L., & Loehlin, J. C. (1976). Resemblance in appearance and the equal environments assumption in twin studies of personality traits. *Behavior Genetics*, *6*(1), 43–52. https://doi.org/10.1007/BF01065677

Polderman, T. J. C., Benyamin, B., de Leeuw, C. A., Sullivan, P. F., van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics*, *47*(7), 702–709. https://doi.org/10.1038/ng.3285

Portenoy, R. K., Ganae-Motan, E. D., Allende, S., Yanagihara, R., Shaiova, L., Weinstein, S., … Fallon, M. T. (2012). Nabiximols for opioid-treated cancer patients with poorly-controlled chronic pain: A randomized, placebo-controlled, graded-dose trial. *Journal of Pain*, *13*(5), 438–449. https://doi.org/10.1016/j.jpain.2012.01.003

Purcell, S. (2013). Statistical Methods in Behavioral Genetics. In R. Plomin, J. C. DeFries, V. S. Knopik, & J. M. Neiderhiser (Eds.), *Behavioral Genetics* (6th ed., p. Appendix). New York: Worth Publishers.

Putnam, F. W. (2018). Ten-Year Research Update Review: Child Sexual Abuse. *Journal of the American Academy of Child & Adolescent Psychiatry*, *42*(3), 269–278. https://doi.org/10.1097/00004583-200303000-00006

R Core Team. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ressler, K. J., & Nemeroff, C. B. (2000). Role of serotonergic and noradrenergic systems in the pathophysiology of depression and anxiety disorders. *Depression and Anxiety*, *12*(S1), 2–19. https://doi.org/10.1002/1520-6394(2000)12:1+<2::AID-DA2>3.0.CO;2-4

Rhee, S. H., Hewitt, J. K., Lessem, J. M., Stallings, M. C., Corley, R. P., & Neale, M. C. (2004). The Validity of the Neale and Kendler Model-Fitting Approach in Examining the Etiology of Comorbidity. *Behavior Genetics*, *34*(3), 251–265. https://doi.org/10.1023/B:BEGE.0000017871.87431.2a

Richardson, K., & Norgate, S. (2005). The equal environments assumption of classical twin studies may not hold. *British Journal of Education Psychology*, *74*(3), 339–350. https://doi.org/10.1348/000709904X24690

Rijsdijk, F. V., & Sham, P. C. (2002). Analytic approaches to twin data using structural equation models. *Briefings in Bioinformatics*, *3*(2), 119–133. https://doi.org/10.1093/bib/3.2.119

Rutter, M., Moffitt, T. E., & Caspi, A. (2006). Gene–environment interplay and psychopathology: multiple varieties but real effects. *Journal of Child Psychology and Psychiatry*, *47*(3–4), 226–261. https://doi.org/10.1111/j.1469-7610.2005.01557.x

Rutter, M., & Quinton, D. (1984). Parental psychiatric disorder: effects on children. *Psychological Medicine*, *14*(4), 853–880.

Rutter, M., & Redshaw, J. (1991). Annotation: Growing up as a Twin: Twin-Singleton Differences in Psychological Development. *Journal of Child Psychology and*

Psychiatry, *32*(6), 885–895. https://doi.org/10.1111/j.1469-7610.1991.tb01916.x

Sadowski, H. S., Ugarte, B., Kolvin, I., Kaplan, C. E., & Barnes, J. (1999). Early life family disadvantages and major depression in adulthood. *British Journal of Psychiatry*, *174*(2), 112–120. https://doi.org/10.1192/bjp.174.2.112

Sartor, C. E., Bucholz, K. K., Nelson, E. C., Madden, P. A. F., Lynskey, M. T., & Heath, A. C. (2011). Reporting bias in the association between age at first alcohol use and heavy episodic drinking. *Alcoholism: Clinical and Experimental Research*, *35*(8), 1418–1425. https://doi.org/10.1111/j.1530-0277.2011.01477.x

Scarr, S., & Carter-Saltzman, L. (1979). Twin method: Defense of a critical assumption. *Behavior Genetics*, *9*(6), 527–542. https://doi.org/10.1007/BF01067349

Schwantes-An, T. H., Zhang, J., Chen, L. S., Hartz, S. M., Culverhouse, R. C., Chen, X., … Saccone, N. L. (2016). Association of the OPRM1 Variant rs1799971 (A118G) with Non-Specific Liability to Substance Dependence in a Collaborative de novo Meta-Analysis of European-Ancestry Cohorts. *Behavior Genetics*, *46*(2), 151–169. https://doi.org/10.1007/s10519-015-9737-3

Sham, P. (1998). *Statistics in Human Genetics* (1st ed.). London: Arnold.

Sherva, R., Wang, Q., Kranzler, H., Zhao, H., Koesterer, R., Herman, A., … Gelernter, J. (2016). Genome-wide Association Study of Cannabis Dependence Severity, Novel Risk Variants, and Shared Genetic Risks. *JAMA Psychiatry*, *73*(5), 472–480. https://doi.org/10.1001/jamapsychiatry.2016.0036

Shiffman, S. (2009). Ecological momentary assessment (EMA) in studies of substance use. *Psychological Assessment*, *21*(4), 486–497. https://doi.org/10.1037/a0017074

Silins, E., Horwood, L. J., Patton, G. C., Fergusson, D. M., Olsson, C. A., Hutchinson, D. M., … Slade, T. (2014). Young adult sequelae of adolescent cannabis use: An integrative analysis. *The Lancet Psychiatry*, *1*(4), 286–293.

https://doi.org/10.1016/S2215-0366(14)70307-4

Slade, T., & Andrews, G. (2005). Latent structure of depression in a community sample: a taxometric analysis. *Psychol Med*, *35*(4), 489–497. https://doi.org/10.1017/S0033291704003708

Snyder, H. R. (2013). Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: A meta-analysis and review. *Psychological Bulletin*, *139*(1), 81–132. https://doi.org/10.1037/a0028727

Solowij, N. (2002). Cognitive Functioning of Long-term Heavy Cannabis Users Seeking Treatment. *The Jounral of the American Medical Association*, *287*(9), 1123–1131. https://doi.org/10.1001/jama.287.9.1123

Sribney, W. (2005). Likelihood-ratio test after survey/robust ML estimation. Retrieved September 10, 2017, from https://www.stata.com/support/faqs/statistics/likelihood-ratio-test/

StataCorp. (2015). Stata Statistical Software: Release 14. *2015*. College Station, TX: StataCorp LP. https://doi.org/10.2307/2234838

Stringer, S., Minică, C. C., Verweij, K. J. H., Mbarek, H., Bernard, M., Derringer, J., … Vink, J. M. (2016). Genome-wide association study of lifetime cannabis use based on a large meta-analytic sample of 32 330 subjects from the International Cannabis Consortium. *Translational Psychiatry*, *6*, e769. https://doi.org/10.1038/tp.2016.36

Sullivan, P. F., Kendler, K. S., & Neale, M. C. (2003). Schizophrenia as a Complex Trait Evidence From a Meta-analysis of Twin Studies. *Arch Gen Psychiatry*, *60*(12), 1187–1192. https://doi.org/10.1001/archpsyc.60.12.1187

Sullivan, P. F., Neale, M. C., & Kendler, K. S. (2000). Reviews and Overviews Genetic Epidemiology of Major Depression : Review and Meta-Analysis. *JAMA*, *287*(9), 1552–1562. https://doi.org/10.1176/appi.ajp.157.10.1552

Swendsen, J. D., Conway, K. P., Degenhardt, L., Glantz, M., Jin, R., Merikangas, K. R., … Kessler, R. C. (2010). Mental disorders as risk factors for substance use, abuse and dependence: Results from the 10-year follow-up of the National Comorbidity Survey. *Addiction*, *105*(6), 1117–1128. https://doi.org/10.1111/j.1360-0443.2010.02902.x

Swendsen, J. D., & Merikangas, K. R. (2000). The comorbidity of depression and substance use disorders. *Clinical Psychology Review*, *20*(2), 173–189. https://doi.org/10.1016/S0272-7358(99)00026-4

Torrens, M., Mestre-Pintó, J.-I., & Domingo-Salvany, A. (2015). *Comorbidity of substance use and mental disorders in Europe*. *European Monitioring Centre for Drugs and Drug Addiction*. https://doi.org/10.2810/532790

Tramèr, M. R., Carroll, D., Campbell, F. A., Reynolds, D. J., Moore, R. A., & McQuay, H. J. (2001). Cannabinoids for control of chemotherapy induced nausea and vomiting: quantitative systematic review. *BMJ (Clinical Research Ed.)*, *323*(7303), 16–19. https://doi.org/10.1136/bmj.323.7303.16

Trucco, E. M., Villafuerte, S., Hussong, A., Burmeister, M., & Zucker, R. A. (2018). Biological underpinnings of an internalizing pathway to alcohol, cigarette, and marijuana use. *Journal of Abnormal Psychology*, *127*(1), 79–91. https://doi.org/10.1037/abn0000310

Turna, J., Patterson, B., & Van Ameringen, M. (2017). Is cannabis treatment for anxiety, mood, and related disorders ready for prime time? *Depression and Anxiety*. https://doi.org/10.1002/da.22664

Van Laar, M., Van Dorsselaer, S., Monshouwer, K., & De Graaf, R. (2007). Does cannabis use predict the first incidence of mood and anxiety disorders in the adult population? *Addiction*, *102*(8), 1251–1260. https://doi.org/10.1111/j.1360-0443.2007.01875.x

Vanyukov, M. M., Tarter, R. E., Kirillova, G. P., Kirisci, L., Reynolds, M. D., Kreek, M.

J., … Ridenour, T. A. (2012). Common liability to addiction and "gateway hypothesis": theoretical, empirical and evolutionary perspective. *Drug and Alcohol Dependence*, *123*(S1), S3–S17. https://doi.org/10.1016/j.drugalcdep.2011.12.018

Verweij, K. J. H., Agrawal, A., Nat, N. O., Creemers, H. E., Huizink, a C., Martin, N. G., & Lynskey, M. T. (2013). A genetic perspective on the proposed inclusion of cannabis withdrawal in DSM-5. *Psychological Medicine*, *43*(8), 1713–1722. https://doi.org/10.1017/S0033291712002735

Verweij, K. J. H., Zietsch, B. P., Lynskey, M. T., Medland, S. E., Neale, M. C., Martin, N. G., … Vink, J. M. (2010). Genetic and environmental influences on cannabis use initiation and problematic use: a meta-analysis of twin studies. *Addiction*, *105*(3), 417–430. https://doi.org/10.1111/j.1360-0443.2009.02831.x.

Wagner, F. A., & Anthony, J. C. (2002). From first drug use to drug dependence: Developmental periods of risk for dependence upon marijuana, cocaine, and alcohol. *Neuropsychopharmacology*, *26*(4), 479–488. https://doi.org/10.1016/S0893-133X(01)00367-0

Ware, M. A., Wang, T., Shapiro, S., Robinson, A., Ducruet, T., Huynh, T., … Collet, J. P. (2010). Smoked cannabis for chronic neuropathic pain: A randomized controlled trial. *CMAJ*, *182*(14). https://doi.org/10.1503/cmaj.091414

Whiteford, H. A., Degenhardt, L., Rehm, J., Baxter, A. J., Ferrari, A. J., Erskine, H. E., … Vos, T. (2013). Global burden of disease attributable to mental and substance use disorders: Findings from the Global Burden of Disease Study 2010. *The Lancet*, *382*(9904), 1575–1586. https://doi.org/10.1016/S0140-6736(13)61611-6

Whiting, P. F., Wolff, R. F., Deshpande, S., Di Nisio, M., Duffy, S., Hernandez, A. V, … Kleijnen, J. (2015). Cannabinoids for Medical Use: A Systematic Review and Meta-analysis. *Jama*, *313*(24), 2456–2473. https://doi.org/10.1001/jama.2015.6358

WHO. (2017). Depression. Retrieved March 11, 2018, from http://www.who.int/mediacentre/factsheets/fs369/en/

Wilkinson, S. T., Yarnell, S., Radhakrishnan, R., Ball, S. A., & D'Souza, D. C. (2016). Marijuana Legalization: Impact on Physicians and Public Health. *Annual Review of Medicine*, *67*(1), 453–466. https://doi.org/10.1146/annurev-med-050214-013454

Williams, R. L. (2000). A note on robust variance estimation for cluster-correlated data. *Biometrics*, *56*(2), 645–646. https://doi.org/10.1111/j.0006-341X.2000.00645.x

World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines*. Geneva: World Health Organization.

Wray, N. R., Lee, S. H., Mehta, D., Vinkhuyzen, A. A. E., Dudbridge, F., & Middeldorp, C. M. (2014). Research Review: Polygenic methods and their application to psychiatric traits. *Journal of Child Psychology and Psychiatry*, *55*(10), 1068–1087. https://doi.org/10.1111/jcpp.12295

Wray, N. R., & Sullivan, P. F. (2017). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *bioRxiv*. https://doi.org/https://doi.org/10.1101/167577

Wykes, T., Haro, J. M., Belli, S. R., Obradors-Tarragó, C., Arango, C., Ayuso-Mateos, J. L., … Wittchen, H. U. (2015). Mental health research priorities for Europe. *The Lancet Psychiatry*, *2*(11), 1036–1042. https://doi.org/10.1016/S2215-0366(15)00332-6

Xian, H., Scherrer, J. F., Eisen, S. A., True, W. R., Heath, A. C., Goldberg, J., … Tsuang, M. T. (2000). Self-Reported Zygosity and the Equal-Environments Assumption for Psychiatric Disorders in the Vietnam Era Twin Registry. *Behavior Genetics*, *30*(4), 303–310. https://doi.org/10.1023/A:1026549417364

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., … Montgomery, G. W. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, *42*(7), 565–569.

Yucel, M., Solowij, N., Respondek, C., Whittle, S., Fornito, A., Pantelis, C., & Lubman, D. I. (2008). Regional brain abnormalities associated with long-term heavy cannabis use. *Archives of General Psychiatry*, *65*(6), 694–701. https://doi.org/10.1001/archpsyc.65.6.694